© 2012 Sreeram Kannan

LAYERING PRINCIPLES FOR WIRELESS NETWORKS

BY

SREERAM KANNAN

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2012

Urbana, Illinois

Doctoral Committee:

Professor Pramod Viswanath, Chair
Associate Professor Chandra Chekuri
Professor Bruce Hajek
Professor Emeritus P. R. Kumar
Professor Muriel Medard, MIT
Professor Rayadurgam Srikant

ABSTRACT

The need to deliver increasing amounts of data over a fixed bandwidth in wireless networks necessitates the engineering of communication schemes with ever-increasing efficiencies. In this context, it is imperative to obtain a fundamental understanding of wireless networks and the engineering architectures that can achieve optimal performance. Accordingly, understanding the capacity, i.e., the maximum information rate supported by wireless networks, is a central aim in the field of information theory. This understanding has been successfully obtained for the case of small canonical networks (called channels, for example, broadcast, multiple-access and interference channels) and for networks under restricted traffic models (for example, unicast, multicast and broadcast). In this thesis, the main goal is to understand the capacity of general wireless networks under a general traffic model, i.e., multiple-unicast, in which there are multiple sources intending to communicate independent information with multiple destinations.

A classical result in undirected wireline networks is the near optimality of routing (flow) for multiple-unicast traffic: the min-cut upper bound is within a logarithmic factor of the number of sources of the max-flow. This establishes the approximate capacity of multiple-unicast in wireline networks. In this thesis, we "extend" this wireline result to the wireless context.

In order to accomplish this, we propose a new model for wireless networks, namely, the polymatroidal network model. In a standard wireline network, the rate of information flow on each edge is constrained by its capacity; in the polymatroidal network model, the capacity of edges that meet at a node are further jointly constrained by a submodular function. While a max-flow min-cut theorem for unicast traffic is known for polymatroidal networks, multiple-unicast traffic has not been studied prior to this work. A key technical contribution of this thesis is an approximate max-flow min-cut theorem for multiple-unicast in undirected polymatroidal networks (the approximation has a multiplicative logarithmic gap). Our key tools are the formulation and analysis of the dual of the flow relaxations

via continuous extensions of submodular functions, in particular the Lovász extension, and the use of metric embeddings into the real line with low average distortion.

In order to translate these results from polymatroidal networks to wireless networks, we propose a natural layered architecture in which information-theoretic schemes are employed at the local level (termed as local physical-layer schemes) and routing is employed at the global level. We show that the feedback and symmetry inherent in wireless networks plays a crucial role in enabling this separation, by demonstrating that the layered architecture is approximately optimal. This result is formally demonstrated for wireless networks under a variety of channel models for which capacity results are known for the corresponding local physical-layer channels. Thus our main result can be viewed as a *meta-theorem*: if there are “good” physical-layer schemes for a certain channel, then, for multiple-unicast in a network composed of such channels, a layered architecture is approximately cut-set achieving.

Finally, we turn our attention to the more general problem of function computation. In the function computation problem, certain nodes of an undirected graph have access to independent data, while some other nodes of the graph require certain functions of the data; this model is motivated by sensor networks and cloud computing. We study the maximum rates at which function computation is possible on a *capacitated graph*; the capacities on the edges of the graph impose constraints on the communication rate. We consider a simple class of computation strategies based on Steiner-tree packing (so-called *computation trees*), which does not involve block coding and has minimal delay.

With a single terminal requiring function computation, the performance of computation trees is known to be optimal when the underlying graph is itself a directed tree, but can be arbitrarily poor in general directed graphs. Our main result is that computation trees are *near optimal* for several classes of function computation requirements even at *multiple* terminals in *undirected* graphs. The key technical contribution here involves connecting prior work in approximation algorithms for Steiner cuts in undirected graphs to the function computation problem. We also demonstrate a certain “duality” between the function computation problem and a communication problem involving multiple multicasts.

To my father, mother, and sister.
To my teachers, advisors and mentors.
To the great Masters, to whom I can but humbly bow.

ACKNOWLEDGMENTS

I am extremely fortunate to be advised by Prof. Pramod Viswanth. His contagious enthusiasm, meticulous thinking and unwavering focus have been great traits for me to imbibe. His deep intuition about the research problem has defined the course of this thesis. I will treasure the many hours that I spent closely working with him, from which I have learned the pleasure of doing scientific research. He has been a mentor par excellence, consistently placing the student's wellbeing higher than his own. Thanks also to Pramod and his wife Suma for impressing on me the possibility of having an independent outlook on every single aspect of life.

I am deeply grateful to Prof. Chandra Chekuri, who has been my unofficial co-advisor and without whom this thesis would have been impossible. Despite my different background, Chandra was tremendously patient and has guided me through unchartered waters. His sharp insight, clear thinking and ability to identify the hardness of problems have been invaluable in setting the directions for this thesis; I hope I have picked up some of these skills from him. He has been extremely kind to me and has gone out of his way to help me.

I would like to thank my other thesis committee members, Prof. R. Srikant, Prof. Bruce Hajek, Prof. P.R. Kumar and Prof. Muriel Medard (MIT), for agreeing to serve on my committee and for finding time in their schedule for my defense. Their suggestions and comments have influenced the writing of this thesis, for which I am grateful. I am thankful to the National Science Foundation, which funded this thesis research by the NSF grants CCF 1017430 and CNS 0721652.

The Coordinated Science Laboratory provided an intellectually charged atmosphere for research. I have had the opportunity to interact with several excellent faculty during my stay as a student. I am grateful to Prof. Srikant for the opportunity to discuss several research problems with him. I thoroughly enjoyed these discussions and have learned a lot from them and from the courses I took with him. I will always look up to him as a role model. I was also fortunate to attend Prof. Hajek's classes, who is one of the best teachers I have had. I will fondly remember the classes of Prof. Kumar, who always brought out the simplicity of the ideas involved in seemingly complex problems. My special thanks to Prof. Muriel Medard (MIT) for agreeing to serve on my thesis committee on such a short notice and for her valuable advice.

It was my pleasure also to get to know some of the other professors in CSL: Prof. Venu Veeravalli, Prof. Nitin Vaidya, Prof. Todd Coleman and Prof. Negar Kiyavash; their warm smiles made trips down the CSL corridors memorable ones. Special thanks to Prof. Veeravalli for motivating me and my teammates to work on the Qualcomm Cognitive Radio Contest. I will vividly remember the fun-filled nights working on the contest with my teammates Adnan, Jayakrishnan and Sreekanth. We enjoyed these sessions so much that it became a regular feature well beyond the contest.

I spent two excellent summers at Qualcomm Research and one awesome summer at Microsoft Research. I am grateful to my mentors Ahmed Sadek, Thomas Zheng, Victor Chan and Madhu Sudan for guiding me through these summers. My special thanks to Madhu Sudan for I thoroughly enjoyed the time spent discussing with him problems as wide-ranging as linguistics, privacy standards, graph theory, random walks and human experiments.

I would like to thank my lab mate Adnan, who has been a great friend, collaborator and awesome travel companion. His joie de vivre, humor, and love for the mountains have made my student tenure a lot of fun. Thanks also to my other labmate and collaborator Quan, whose unending supply of energy continues to inspire me. I would like to acknowledge my gratitude to my former lab-mates Nilesh and Vinod, whose advice shaped my decision to join UIUC. Thanks are also due for the other CSL students Vineet, Siva Theja and Sachin, who have always been ready for varied discussions on topics across the spectrum. They have also been excellent classmates and project mates in the courses that I was fortunate to take with them. The other denizens of CSL 130, Rui, Juan, Farzaneh, Xun and Dan, have made it a second home for me.

A close circle of friends made the vicissitudes of the Champaign weather and graduate school more tolerable, even enjoyable, by making jokes out of them. My heartfelt thanks to Gayathri, Anand, Jayakrishnan, Srikanthan, Karthik, Hemant, Vijay, Arvind Ganesh, Vivek, Jayanand, Anjan, Rajan and Kunal for the amazing time I had at Champaign. It is hard to imagine my tenure as a student without them. Special thanks to my roommate, Chaitanya, whose limitless hospitality I have always been glad to bask in.

I would also like to thank the administrative staff at CSL and ECE who made life a lot easier for me. Special thanks to Barbara Horner for being the best admin that I have ever met. Her efficiency, warmth and energy radiate into the CSL first floor. Thanks also to Dan Jordan for his help. I would also like to thank Jamie Hutchinson for his careful editing of the thesis.

Before coming to UIUC, I spent two fun-filled years studying at IISc under Prof. Vijay Kumar. These years shaped my interest in information theory and wireless networks; I am extremely grateful to him for introducing me to research. I would also like to sincerely thank Prof. P.V. Ramakrishna, under whom I spent three years working as an undergraduate research associate. He played a crucial role in kindling my curiosity for technology and mathematics in my formative years.

I would like to convey my deep sense of gratitude to my sister Srilakshmi, my brother-in-law Karthik and my niece Varsha for their unflinching love and support; I cherish the several months I got to spend with them during my internships. Finally, I would like to thank my father, Dr. S. Kannan and my mother, Dr. Rukmani for always giving me unconditional support, love and freedom; words are insufficient to convey my gratitude to them.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

"I do not think that the wireless waves I have discovered will have any practical application" – Heinrich Hertz

1.1 Background

Wireless data usage is increasing exponentially, whereas the bandwidth allocated to carry the data grows very slowly. Current predictions suggest that wireless data requirements will increase eighteen-fold in the next five years [29]. A simple, although expensive, way to meet this demand is to increase the bandwidth allocation for wireless telephony correspondingly. However, the bandwidth allocation for cellular telephony is unlikely to more than double in this five-year period. Thus, in order to meet these mounting demands on wireless systems, the future wireless systems need to be designed to deliver ever-increasing throughput over a given bandwidth. This calls for a fundamental understanding of optimal wireless architectures and communication schemes. The main goal of this thesis is to examine this problem from an information theoretic viewpoint.

Wireless networks are typically engineered using a modular approach, commonly referred to as a layered approach. In this approach, the whole task of the network is subdivided into various layers (modules) and each layer can then be essentially independently designed. The layering is broadly as follows: the physical layer deals with the channel uncertainty, the medium access control layer deals with scheduling of users in the wireless context, the network layer handles routing of information and the transport layer deals with network congestion. While this design methodology has several engineering advantages that have led to the proliferation of wireless networks, a fundamental understanding of the capacity of wireless networks[1] and the structure of "optimal" architectures is still lacking.

[1] A wireless network is specified by the connectivity between nodes, the channel model and the traffic model, i.e., which nodes have messages to send, and who wants to receive those messages.

This leads us to the following question.

Question 1: *What is the capacity of wireless networks?*

This is the central question of study in the field of network information theory. While characterization of exact capacity is an unsolved problem for all but the simplest networks, substantial progress has been made in the recent past in understanding the capacity *approximately* of two special classes of wireless networks (by approximate, we mean that the performance of the achievable scheme is guaranteed to be within a certain additive or multiplicative factor of the best possible scheme).

- *Simple wireless networks under general traffic models*: Prior work has characterized the approximate capacity of several simple wireless networks, i.e., wireless networks where all communication happens in one-hop, under the multiple-unicast traffic model. These simple networks are referred to as channels, and approximate capacity characterizations exist for the multiple-access channel [31], broadcast channel [153], 2-user interference channel [40] and the K-user interference channel [21].

- *General wireless networks under simple traffic models*: Prior work has obtained the approximate capacity of unicast, multicast and broadcast (one source conveying independent information to multiple destinations) traffic in general wireless networks [17, 91, 9, 69].

In order to show a capacity result, two main ingredients are necessary: an achievable scheme along with its rate characterization and an outer bound, which upper-bounds the communication rates of any achievable scheme. A well-known upper bound in network information theory is the cut-set upper bound, which can be written down for any problem instance. While this is not necessarily the best known bound even in the special cases described above, the rates suggested by this bound are achievable *approximately*, with a gap that is additive in some instances (unicast in general networks) and multiplicative in certain others (multiple unicast in interference channel).

If one source wants to talk to one (multiple) destination, we call it the unicast (multicast) traffic model. If there are multiple sources each wanting to talk to one (multiple) destination, then the traffic model is called multiple-unicast (multiple-multicast). Multiple-multicast is the most general traffic model.

The capacity region of a wireless network is the set of rate tuples at which the sources can communicate to the corresponding destinations.

Given this state of the art, the exact resolution of Question 1 remains distant. However, we can ask the following questions, which are more modest.

Question $1a$**:** *What is the approximate capacity of multiple-unicast in general wireless networks?*

Question $1b$**:** *Is the cut-set bound approximately achievable for multiple-unicast?*

The main result of this thesis is a successful resolution of Question $1a$ by showing that the cut-set bound is approximately achievable, thus answering Question $1b$ in the affirmative. Our work builds on the extensive prior work conducted in this area, and we provide a detailed survey of these results in the corresponding chapters.

Our main result shows that our proposed scheme achieves the cut-set bound to within a multiplicative gap of $\mathrm{O}(\log \mathrm{k})$ and a constant additive factor (which depends on the particular channel model) in the worst case, where k is the number of unicasts. Ignoring the constant additive factor (which is present even in the unicast case), this worst-case $\mathrm{O}(\log \mathrm{k})$ factor is interesting for the following reasons.

- A well provisioned network should have the total resources scaling with k, and thus the cut-set bound on the sum-capacity of a well-provisioned network will be of order $\Omega(k)$. Our results state that the scheme can achieve a sum-rate $\Omega(\frac{k}{\log k})$ which is reasonably close to k. In contrast to this $\mathrm{O}(\log \mathrm{k})$ characterization, the only known scheme for multiple-unicast in a general wireless network, i.e., time sharing, will achieve a constant sum-rate irrespective of the number of users.

- This is the tightest known characterization of the capacity of even *wireline* networks, which are a special case of the networks considered here.

- This is a tight characterization of our proposed achievable scheme, i.e., there exists wireless networks in which the achievable rate and the cut-set bound will differ by a factor $\mathrm{O}(\log \mathrm{k})$. Furthermore, it is possible that our scheme may be within a constant factor of a more nuanced outer bound. This is related to one of the open questions in wireline network theory.

While existing works attempt to obtain the approximate capacity for specific instances of the problem, we adopt a different viewpoint and focus our attention on obtaining general results for arbitrary networks at the expense of obtaining potentially weaker approximation in specific instances.

While the answer to Question 1 has been known for special traffic models (single-unicast, for example) from previous work, the architectures suggested by these works do not correspond to layered schemes, which are amenable to engineering design. In this work, we take a step further and ask the following question, which is more refined than the first question.

Question 2**:** *Are there simple layered architectures that achieve the capacity approximately?*

In this thesis, we answer this question also in the affirmative. In fact, the way we answer Question 1 is via answering Question 2, which we accomplish by showing a *layered* achievable strategy that can achieve within $\mathrm{O}(\log \mathrm{k})$ of the cut-set bound. The pioneering work [80] has proved that a layered architecture is optimal in the case of a wireline network, which is composed of independent noisy links. This thesis is an attempt to generalize this result to the more general context of wireless networks.

1.1.1 Program outline

A general technique employed in prior work (see, for example, the unicast result in [17]) to show approximate capacity characterizations proceeds in three steps. In this thesis also, we follow the same broad program.

- *Step 1:* Identify the corresponding results for wireline networks (in the unicast case, this is the max-flow min-cut theorem)
- *Step 2:* Extend the results to a generalization of the wireline network (in the unicast case, it is the linear deterministic network)
- *Step 3:* "Lift" the results from the generalized wireline network to the wireless network.

The multiple-unicast problem in general (directed) wireline networks is known however to be hard to solve, the difficulty being two-fold:

1. Flow (routing) is not known to be optimal and network coding (i.e., combining of information from various sources while the information passes through the network) is known to be required. In fact, network coding can give arbitrarily large gains over routing [59].

2. The flow rate-region is not even approximately close to the cut-set bound and furthermore, it is NP-hard to approximate the cut-set bound to within a n^ϵ factor for some $\epsilon > 0$ [28], suggesting that no polynomial time coding scheme can get provably close to the cut-set bound.

Thus it appears that the characterization of the capacity would require both better schemes than routing and better outer bounds than cut-bounds. While there are small problem instances where we know how to obtain both, in the general case, both have proven elusive.

While the picture painted above is gloomy, the following information comes as the silver lining in the cloud: in *undirected* wireline networks, there is no known instance where network coding beats routing (flow) and in fact, flow and cut are within a worst-case factor $\mathrm{O}(\log \mathrm{k})$ of each other. The latter fact is a celebrated result of Leighton and Rao [90], which was later generalized by Linial, London and Rabinovich [97]. Also, since wireless networks have natural symmetry, i.e., when node A talks to node B, node B can talk back to node A, they seem closer to undirected networks than directed networks. Thus we will adopt undirected wireline networks as our starting point.

Key simplification: *Consider* symmetric *wireline and wireless networks.*

The identification of the approximate capacity results for multiple-unicast in undirected wireline networks as a baseline completes the first step of our program.

1.1.2 Polymatroidal networks

Now, we proceed to the second step of our program, which is to generalize this result to a model of intermediate complexity between wireline and wireless networks. Linear deterministic models, proposed in [17], capture the broadcast and superposition nature of the wireless medium while suppressing the effect of noise. However, in the case of multiple unicast traffic, even this model turns out to be too complex to solve. The main reason for this complexity is that even the resolution of the unicast problem in this model requires *network coding*. While the simplest flavor of network coding, namely, random network coding works well in the unicast scenario, it turns out to be a poor strategy when there is more than one source-destination pair in the network; this is because random mixing of packets introduces inter-session interference while routing (flow) keeps the information of different sessions separate. This leads us to the following question.

Question 3**:** *Are there network models that generalize wireline networks and capture key features of wireless networks, and yet are amenable to the study of multiple-unicast traffic?*

We answer this question in the affirmative by proposing an *undirected polymatroidal network model* for wireless networks. Polymatroidal networks are similar to standard wireline networks (with edge capacities); however, in addition to having edge capacity constraints, there are joint capacity constraints on the set of edges which meet at a vertex. These joint constraints are characteristic of wireless networks, where transmission on one link constrains the rate of transmission on another. Directed polymatroidal networks were introduced in [88, 60] and they were applied in the information theory literature [146] in the context of broadcasting in (directed) deterministic networks. Undirected polymatroidal networks, introduced in this thesis, have the further nice property that they generalize edge-capacitated and node-capacitated networks simultaneously, thus unifying their treatment.

A key technical contribution of this thesis is to extend the work of Leighton and Rao [90], which showed that routing is near-optimal in undirected wireline networks, to undirected polymatroidal networks, thus showing that routing can get to within a factor of $\mathrm{O}(\log \mathrm{k})$ of the cut-set bound in the case of multiple-unicast traffic. This result is established by resorting to continuous extensions (Lovász extension and convex closure) of the submodular functions encountered in polymatroidal networks and using the technique of embedding metric spaces into the real line with small average distortion, introduced by Rabinovich [123].

While the worst-case logarithmic flow-cut gap is the best possible in general wireline networks, there may be some special families of networks where the gap can be reduced. This leads us to the following question.

Question 4**:** *Can we obtain better than logarithmic flow-cut gaps for special instances of multiple-unicast in polymatroidal networks?*

We answer this question in the affirmative by showing several examples where sharper results for wireline networks can be generalized to polymatroidal networks. In particular, we show that there is a constant factor flow-cut gap for *sum-rate* in the following cases.

- There are M sources and N destinations and there are messages to be sent from each source to each destination. The constant factor in this case is 2.
- There is a group of G nodes, wanting to communicate messages to all other

nodes in the group. The constant factor in this is also 2.

- The communication graph is planar.

While the first two results follow quite easily from the proof for the general case, the result for planar graphs requires additional care and attention.

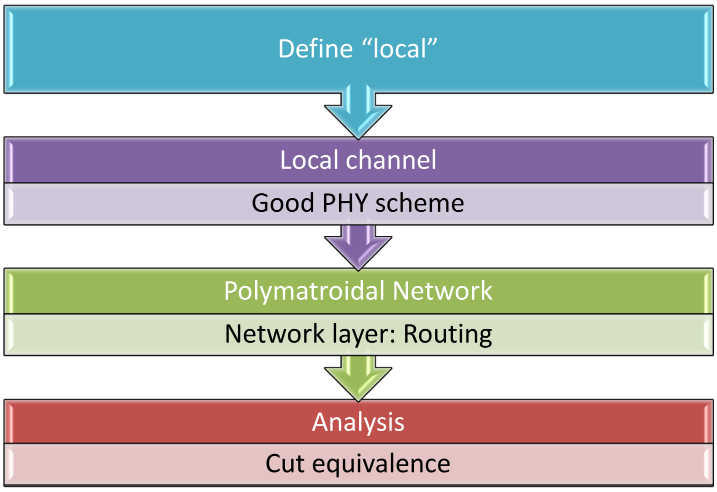


Figure 1.1: Layered architecture

Thus, having accomplished the second step of our program (obtain models for wireless networks), we can proceed to the third step, which is to lift the results from polymatroidal to wireless networks. This is done by using a layered architecture (illustrated in Fig. 1.1) as follows:

- The network is decomposed into a set of local physical layer channels.

- Information-theoretic coding schemes are applied to these local channels. Thus each one of these channels can be replaced by their corresponding capacity regions. If the communication schemes achieve close to the cut-set bound for these channels, then the rate region is polymatroidal. In the case of general wireless networks, the codign schemes for the physical layer channels are based on interference alignment [21].

- In this obtained polymatroidal network, routing achieves close to the cut-set bound.

- By demonstrating that cut-set bounds in the polymatroidal and the wireless network are close, we obtain the necessary result.

This establishes the requisite approximate capacity characterization of wireless networks, which we accomplish for various channel models in this thesis.

1.1.3 Function computation

Having established these promising results in the context of communicating information in wireless networks, it should be noted that there are other scenarios in networks that do not fit the paradigm of communication. In particular, in sensor networks, the sensor fusion node may be interested in the computation of a function of the sensor readings rather than in obtaining the value of each of these sensors. Similarly, in a cloud computing scenario, the data may be stored in a distributed manner, with different types of information about the same record stored in different locations, and one may be interested in computing a function of the data. The class of problems that arises in this setting is studied under the general area of *in-network function computation*, where nodes in the network perform intermediate computations so that the end-to-end goal of computation (rather than communication) is realized at maximum possible rate. Our success in the communication problem immediately prompts us to ask the following question.

Question 5**:** *Can approximate capacity characterizations be obtained in the context of function computation?*

In this thesis, we answer this question in the affirmative in the following context. Consider a *wireline* edge-capacitated network where there are multiple source nodes (think of them as sensor nodes) and a single sink node that wishes to compute a function (for example, the arithmetic sum) of the data at the sources. We are interested in the rate at which such a computation can be performed. A generalization of this setting is when there are multiple sets of sensors, each one corresponding to a different modality like temperature, pressure, etc., which share the common communication infrastructure. We call this setting the multi-session function computation problem. In this case, there are k sessions, each comprising t sensors and a fusion center, and the fusion center wants a certain function of its group of t sensors. We refer to this as multi-session function computation.

The single-session function computation problem was originally formulated by Giridhar and Kumar [51] and this work was built upon significantly in [82, 12].

The multi-session function computation problem has received relatively little attention, some exceptions being [128, 87, 126]; in these works, each destination demands the sum of *all* the sources. This setup is different from the setting in the current thesis, where each destination demands *general functions* of *distinct variables*.

The cut-set bound can be generalized to the function computation setting to obtain an outer bound. We would like to ask if there is a natural class of simple schemes that can achieve the cut-set bound approximately. While it is well known that routing is far from optimal in this setting, there is another class of schemes called *computation trees*, which has been used in [51, 82, 12, 133], where the function is computed by passing information along a Steiner tree with in-network computation performed at junction nodes and the other nodes simply forwarding the information. This class of strategies is related to Steiner packings, which are commonly used for multi-casting in wireline networks. This leads us to the following question.

Question 6**:** *Can a connection between (multiple) multi-casting and (multi-session) function computation be formalized?*

We answer this question in the affirmative by showing that this is indeed true when the function to be computed is linear. To obtain this duality we just reverse the nature of the sources and the destination; thus the fusion center transmits data that needs to be received by all the sensor nodes. Using this duality, we can easily show that for single-session linear-function computation, computation trees are within a factor of 2 of the cut-set bound. This settles the question posed in [133] about the near-optimality of Steiner tree packing in this setting. Furthermore, in the case of multi-session linear-function computation, computation trees are shown to be within a factor of $\mathrm{O}(\log(\mathrm{kt}))$ of the cut-set bound which is the dual for a similar result for multiple multicast in wireline networks. The factor gap increases in the case of other more complex functions, but we provide an upper-bound for this gap in this thesis.

1.2 Summary of Results

An important contribution of this thesis is the following *meta-theorem*:

> **Meta Theorem:** *If there is a "good" physical-layer scheme (which is approximately cut-set achieving and reciprocal) for a certain channel,*

then, for multiple-unicast in a network composed of such channels, a layered architecture is approximately cut-set achieving (to within a logarithmic factor in the number of messages).

The following is a summary of the key ideas which are used in this paper to argue that a layered architecture is approximately capacity optimal:

- Model a wireless network as a bidirected network, by using the natural reciprocity of wireless networks.
- Utilize a good local "physical layer" scheme for each channel and identify the combinatorial structure of the rate region (typically submodularity).
- Show that local physical layer schemes convert a wireless network into a *bidirected polymatroidal network*. Thus the bidirected polymatroidal network can be viewed as a *graphical model* for wireless networks.
- Prove a Leighton-Rao type approximation result for bidirected polymatroidal network, which shows that routing is near optimal for k-unicast traffic.
- Argue that the layered architecture with local physical layer scheme + global routing achieves the cut-set approximately in the wireless network.
- We provide a technique by which "good" results for a given channel can be *lifted up* to good results for a general network comprised of those channels.

We justify the meta-theorem formally in the context of the following channel models:

1. Networks composed of Gaussian broadcast and MAC channels
2. Networks composed of broadcast erasure channels with feedback
3. Fast fading wireless networks
4. Degrees-of-freedom approximation for fixed wireless networks
5. Linear deterministic networks composed of MAC and broadcast channels
6. Networks composed of MIMO MAC and broadcast channels with delayed CSI feedback

7. Fast fading linear deterministic networks

For each of these networks, under a general k-unicast traffic model, the approximation factor on the rate is $\mathrm{O(\log k)}$ for the *entire rate region* in addition to the loss incurred due to the physical layer scheme, which is typically a power-scaling loss. Under more specific traffic models, such as the X-traffic model (where each of the J sources have messages to send to each of the K destinations) or a "group-communication" traffic model (where a subset S of nodes have messages to send to each other), we prove a constant approximation factor for the *sum-rate*, again in addition to a power scaling loss (the constant being 1, 2 or 4 depending on the specific channel and traffic model).

Furthermore, we show that similar results hold even in the context of function computation in wireline networks. The key technical contribution there involves connecting existing approximation algorithms for Steiner cuts in undirected graphs to the function computation problem.

1.3 Organization

The rest of this thesis is organized into various chapters as follows. While there is a continuity in these chapters, the thesis is written in such a way that each chapter can be read in a self-consistent manner.

- Polymatroidal networks are motivated, defined and studied in Chapter 2. While the main result of this chapter is an approximate max-flow min-cut theorem for multiple-unicast traffic in undirected polymatroidal networks, there are several other results on polymatroidal networks throughout this chapter.

- In Chapter 3, we consider multiple-unicast traffic in wireless networks. We study wireless networks under various channel models and either study existing physical layer schemes or construct new ones so that a layered architecture based on these physical layer schemes converts the wireless network into a polymatroidal network. We finally show how to translate outer-bounds from the polymatroidal network to the wireless network, which results in approximate capacity results for wireless networks.

- Then, we study function computation in capacitated graphs in Chapter 4, where we show that a simple achievable strategy based on tree packing can get close to the cut-set bound for certain function classes.
- We conclude in Chapter 5 by stating the open questions that arise from this research.

CHAPTER 2

POLYMATROIDAL NETWORK MODEL

> "There is no practical question on which anything more than an approximate solution can be had." – Ralph Waldo Emerson

In this chapter, we will generalize results for multiple unicast in wireline networks to a class of networks called polymatroidal networks. As pointed out earlier, for undirected wireline networks, it is known [90, 97] that the flow rate region (routing) and cut-set bound (outer bounds on any scheme) are within a factor of $O(\log k)$ of each other. In this chapter, we show that this result generalizes to polymatroidal networks as well.

2.1 Problem Setup

Consider a communication network represented by a directed graph $G = (V, E)$. In the so-called edge-capacitated scenario, each edge e has an associated capacity $c(e)$ that limits the information flowing on it. We consider a more general network model called the *polymatroidal network model* introduced by Lawler and Martel [88] and independently by Hassin [60]. This model is closely related to the submodular flow model introduced by Edmonds and Giles [37]. Both models capture as special cases, single-commodity s-t flows in edge-capacitated directed networks, and polymatroid intersection, hence their importance. Moreover the models are known to be equivalent (see Chapter 60 in [131], in particular Section 60.3b). The polymatroidal network flow model is more directly and intuitively related to standard network flows and one can easily generalize it to the multi-commodity setting which is the focus of this chapter.

The polymatroidal network flow model differs from the standard network flow model in the following way. Consider a node v in a directed graph G and let $\delta^-_G(v)$ be the set of edges into v and $\delta^+_G(v)$ be the set of edges out of v. In the standard model each edge (u, v) has a non-negative capacity $c(u, v)$ that is independent of other edges. In the polymatroidal network for each node v there are two associated submodular functions (in fact polymatroids[1]) ρ_v^- and ρ_v^+ which impose joint capacity constraints on the edges in $\delta_G^-(v)$ and $\delta_G^+(v)$ respectively. That is, for any set of edges $S \subseteq \delta_G^-(v)$, the total capacity available on the edges in S is constrained to be at most $\rho_v^-(S)$, similarly for $\delta_G^+(v)$. Note that an edge (u, v) is influenced by ρ_u^+ and ρ_v^-. Lawler and Martel considered the problem of finding a maximum s-t flow in this model. The results in [88, 60] show that various important properties that hold for s-t flows in standard networks generalize to polymatroid networks; these include the classical maxflow-mincut theorem of Ford and Fulkerson (and Menger) and the existence of an integer valued maximum flow when capacities are integral.

2.1.1 Motivation

In this thesis, the main motivation to study polymatroidal networks stems from their ability to model wireless networks. A node in a wireless network communicates with several nodes over a broadcast medium and hence the channels interfere with each other; this imposes joint capacity constraints on the channels. Thus the interference links in the original wireless network are replaced by bit-pipes, the rates of which are constrained to lie in the capacity region of the corresponding interference channel. Since several interference scenarios of interest correspond to (almost) polymatroidal capacity regions, the polymatroidal network model serves as a proxy for the wireless network. Any communication scheme in the polymatroidal network corresponds to a "layered" scheme in the wireless network, i.e., one in which each channel is operated using an information theoretic coding scheme, and then the scheme for the polymatroidal network is run on top of that.

The original motivation for the Lawler-Martel polymatroidal network model came from an application to a scheduling problem [102]. This model has other applications as well; for example, in transportation networks, the edges normally represent transportation resources (roads or railway lines) and nodes model important junction points. In standard wireline networks, the maximum flow on a given edge is modeled using a capacity function and the goal is to study schemes that

[1] A set function $f : 2^N \to \mathbb{R}$ over a finite ground set N is submodular iff $f(A) + f(B) \geq f(A \cap B) + f(A \cup B)$ for all $A, B \subseteq N$; equivalently $f(A \cup \{i\}) - f(A) \geq f(B \cup \{i\}) - f(B)$ for all $A \subset B$ and $i \notin A$. It is monotone if $f(A) \leq f(B)$ for all $A \subset B$. In this thesis, a polymatroid refers to a non-negative monotone submodular function with $f(\emptyset) = 0$.

achieve maximum flow from a given source node to a destination node. However, in several cases, the maximum rate of traffic flow on an edge can depend also on the flow of traffic on other edges incident at a node; this can be modeled using a submodular cost function leading to a polymatroidal network model. The dual of this max-flow problem in the standard wireline network is the min-cut problem, which, in the transportation network example, captures the minimum number of railway lines that need to be cut in order for the source and destination to be unable to communicate with each other. In this case, it may make sense to prefer cuts (set of edges) for which many railway lines pass through the same city as opposed to cuts involving railway lines going through distinct cities, even if they cut the same number of railway lines. In other words, the value of the cut is not simply the sum of the number of railway lines cut, but depends on whether edges belonging to the same node are cut or not. This is modeled neatly using a submodular cost function as well.

More recently, there have been several applications of a related network model called the linking systems model [132], to information flow in wireless networks [17, 8, 160, 52, 124, 71]. However, these results are fundamentally different on two counts: the linking-systems framework does not serve as a proxy for wireless networks, i.e., schemes for the linking systems model do not have any corresponding schemes in the wireless networks (although this defect was partially remedied in [124]). Furthermore, only single-commodity flows (single-unicast problems) have been studied in this model and multi-commodity flow under this model remains unexplored.

Most of the work on polymatroidal networks so far has focused on the case of a single unicast (i.e., there is only one $s - t$ pair). In this thesis, we consider the scenario where several source-sink pairs $(s_1, t_1), (s_2, t_2), \ldots, (s_k, t_k)$ share the capacity of the network. This is referred to as the multiple unicast setting in the communication literature and as *multicommodity* flows and cuts in the theoretical computer science literature. While the primary motivation is applications to (wireless) network information flow, another motivation is to understand the extent to which techniques and results that were developed for multicommodity flows and cuts in standard networks generalize to polymatroidal networks. We note that polymatroidal networks allow for a common treatment of edge and node capacities; an advantage is that one can define cuts with respect to edge removals while the cost is based on nodes. As far as we are aware, multicommodity flows and cuts in polymatroidal networks have not been studied previously.

Flow-cut gaps in polymatroidal networks: The main focus of this chapter is understanding multicommodity flow-cut gaps in polymatroidal networks. In communication networks cuts can be used to information theoretically upper-bound achievable rates while flows allow one to develop lower bounds on achievable rates by combining a variety of routing and coding schemes. Flow-cut gaps are of therefore of much interest. Unlike the case of single-commodity flows where maximum flow is equal to minimum cut, it is well-known that even in standard edge-capacitated networks no tight min-max result holds when the number of source-sink pairs is three or more (two or more in case of directed graphs). See [131] for some special cases where min-max results do hold. Flow-cut gap results have been extensively studied in theoretical computer science starting with the seminal work of Leighton and Rao [90]. The initial motivation was approximation algorithms for cut and separator problems that are NP-hard. There has been much subsequent work with a tight bound of $O(\log k)$ established for flow-cut gaps in undirected graphs in a variety of settings [48, 97, 16, 45]. It has also been shown that strong lower bounds exist for flow-cut gaps in directed graphs; for instance the gap is $O(\min\{k, n^{\delta}\})$ between the maximum concurrent flow and the sparsest cut [130, 28] where δ is a fixed constant. However, poly-logarithmic upper bounds on the gaps are known for the case of symmetric demands in directed graphs [77, 41]. Motivated by the above positive and negative results we focus on those cases where poly-logarithmic flow-cut gaps have been established. We show that several of these gap results extend to polymatroid networks. Our results and techniques lead to new approximation algorithms for cut problems in polymatroidal networks which could have future applications. However, in this thesis we restrict our attention to quantifying flow-cut gaps.

Bidirected and undirected polymatroidal networks: As we mentioned already, strong lower bounds exist on flow-cut gaps for directed networks. Positive results in the form of poly-logarithmic upper bounds on flow-cut gaps for standard networks hold when the demands are symmetric or when the supply graph is undirected. A natural model for wireless networks is the *bidirected* polymatroidal network. For two nodes u and v in a wireless network, it is a reasonable approximation to assume that the channel from u to v is similar to that from v to u; hence one can assume that the underlying graph G is bidirected in that if the edge (u, v) is present then so is (v, u). Moreover, we assume that for any node v and $S \subseteq \delta^-(v)$, $\rho_v^-(S) = \rho_v^+(S')$ where $S' \subseteq \delta^+(v)$ is the set of edges that correspond to the reverse of the edges in S. Within a factor of 2, bidirected polymatroidal networks can be approximated by *undirected* polymatroidal networks: we have an undirected graph G and for each node v a single polymatroid ρ_v that constrains the capacity of the edges $\delta_G(v)$, the set of edges incident to v. The main advantage of undirected polymatroid networks is that we can use existing tools and ideas from metric embeddings to understand flow-cut gap results. Undirected polymatroidal networks have not been considered previously. We observe that they allow a natural way to capture both edge and node-capacitated flows in undirected graphs. To capture node-capacitated flows[2] we set $\rho_v(S) = 2c(v)$ for all $\emptyset \neq S \subseteq \delta(v)$ where $c(v)$ is the capacity of v. We mention an advantage of using polymatroidal networks even when considering the special case of node-capacitated flows and cuts: one can define cuts with respect to edges even though the cost is on the nodes. This is in fact quite natural and simplifies certain aspects of the algorithms in [45].

2.1.2 Overview of results and technical ideas

We do a systematic study of flow-cut gaps in multicommodity polymatroidal networks, both directed and undirected. Let $G = (V, E)$ be a polymatroidal network on n nodes with k source-sink pairs $(s_1, t_1), \ldots, (s_k, t_k)$. We consider two flow problems and their corresponding cut problems: (i) maximum throughput flow and multicut (ii) maximum concurrent flow and sparsest cut.

Our main results are tabulated in Table 2.1 and are summarized below.

- For directed networks we show a reduction based on the dual that establishes a correspondence between flow-cut gaps in polymatroidal networks and the standard edge-capacitated networks. This allows us to obtain polylogarithmic upper bounds for flow-cut gaps in directed polymatroidal networks with *symmetric* demands via results in [77, 41] for both throughput flow and concurrent flow. In particular we obtain an $O(\min\{\log^3 k, \log^2 n \log\log n\})$ gap between the maximum concurrent flow and sparsest cut. The reduction is applicable only to directed graphs.

[2] The factor of 2 is needed since a flow path p through an internal node v uses two edges. On the other hand, it is not needed for the sources and sinks. This technical issue is a minor inconvenience with undirected polymatroidal networks; we note that this also arises in treating node-capacitated multicommodity flows [45].

Table 2.1: Summary of Results

Setting	Max. Concurrent Flow / Sparsest Cut Gap	Max. Throughput Flow / Multicut Gap
Undirected polymatroidal network	$O(\log k)$	$O(\log k)$
Directed polymatroidal network (symmetric demands)	$O(\min\{\log^2 k, \log n \log\log n\})$	$O(\min\{\log^3 k, \log^2 n \log\log n\})$
Planar undirected polymatroidal network	[Open]	$O(1)$

- We show that line embeddings with low average distortion [104, 123] lead to upper bounds on flow-cut gaps in polymatroidal networks — this connection is inspired by the work in [45] for node-capacitated flows. For undirected polymatroidal networks this leads to an optimal $O(\log k)$ gap between maximum concurrent flow and sparsest cut. We also obtain an optimal $O(\log k)$ gap between throughput flow and multicut. These imply corresponding results for bidirected networks. Furthermore, similar to [45], we exploit the embedding connection to obtain improved $O(\sqrt{\log k})$-approximation algorithms for sparsest cut problems with product demands using stronger relaxations via semi-definite programming (and associated embedding theorems) [14, 2, 15].

- We consider polymatroidal networks that exclude a fixed graph K_h as minor (this includes planar graphs). We show an $O(h^2)$ gap between the maximum throughput flow and minimum multicut for these networks. As a corollary, we obtain a constant factor approximation for *node-weighted* multicut in such graphs. Our result is based on a reinterpretation of the chopping operation in the network decomposition theorem in [76] as a line embedding. It has been conjectured [57] that there is a corresponding constant gap result for the maximum concurrent flow problem in the case of standard planar (and more generally, minor-free) networks; however this conjecture remains unresolved.

Most of the literature on multicommodity flow-cut gaps is based on analyzing the dual of the linear program for the flow which can be viewed as a fractional relaxation for the corresponding cut problem. The gap is established by showing the existence of an integral cut within some factor of the relaxation. For standard edge and node-capacitated network flows the dual linear program has length variables on the edges which induce distances on the nodes. The situation is more involved in polymatroidal networks, in particular, the definition of the cost of a cut is somewhat complex and is discussed in more detail in Section 2.2.2.

Our starting point is the use of the Lovász extension of a submodular function [98] to cleanly rewrite the dual of the flow linear programs. This simplifies the constraint structure of the dual at the expense of making the objective a convex function. However, we are able to exploit properties of the Lovász extension in several ways to obtain our results. Our techniques give two new dual-based proofs of the maxflow-mincut theorem for single commodity polymatroid networks that was first established by Lawler and Martel algorithmically [88] via an augmenting path based approach. We believe that the applicability of embedding based methods for polymatroidal networks is of independent mathematical interest.

For the most part we ignore algorithmic issues in this thesis although all the flow-cut gap results lead to efficient algorithms for finding approximate cuts.

2.1.3 Organization

The rest of this chapter is organized as follows. Two different ways of formulating multicommodity flows and cuts are defined formally for polymatroidal networks in Sec. 2.2. In Sec. 2.3, we develop several convex programming relaxations for the combinatorial problem for minimum cut. This relaxation is exploited in Sec. 2.4 to show logarithmic flow-cut gaps for directed polymatroidal networks under "symmetric demands" by using a reduction from the polymatroidal network problem to the standard network problem. In Sec. 2.5, logarithmic flow-cut gaps are shown for undirected polymatroidal networks. Finally, for the special case of planar and minor-free graphs, we develop stronger flow-cut gaps in Sec. 2.6.

2.2 Multicommodity Flows and Cuts in Polymatroidal Networks

We let $G = (V, E)$ represent a graph whether directed or undirected. We use (u, v) for an ordered pair of nodes and uv to denote an unordered pair. In a directed graph G, for a given node v, $\delta^-_G(v)$ and $\delta^+_G(v)$ denote the set of incoming and outgoing edges at v. In undirected graphs we use $\delta_G(v)$ to denote the set of edges incident to v. We omit the subscript G if it is clear from the context. We are interested in multicommodity flows and cuts. In addition to the graph, the input consists of a set of k source-sink pairs $(s_1, t_1), \ldots, (s_k, t_k)$ that wish to communicate independently and share the network capacity.

In a directed polymatroidal network, each node $v \in V$ has two associated polymatroids ρ^-_v and ρ^+_v with ground sets as $\delta^-(v)$ and $\delta^+(v)$ respectively. These functions constrain the joint capacity on the edges incident to v as follows. If $S \subseteq \delta^-(v)$, then $\rho^-_v(S)$ upper-bounds the total capacity of the edges in S; similarly, if $S \subseteq \delta^+(v)$, then $\rho^+_v(S)$ upper-bounds the total capacity of the edges in S. We assume that the functions $\rho^-_v(\cdot), \rho^+_v(\cdot), \quad v \in V$, are provided via value oracles. In undirected polymatroidal graphs we have a single function $\rho_v(\cdot)$ at a node v that constrains the capacity of the edges incident to v. Continuous extensions of submodular functions, namely the Lovász extension [98] and the convex closure, are important technical tools in interpreting and analyzing the duals of the linear programs for multicommodity flow in the polymatroid setting. We discuss these in Section 2.2.2. We first discuss the two flow problems of interest, namely maximum throughput flow and the maximum concurrent flow.

2.2.1 Flows

A multicommodity flow for a given collection of k source-sink pairs $(s_1, t_1), \ldots,$ (s_k, t_k) consists of k separate single-commodity flows, one for each pair (s_i, t_i). The flow for the i'th commodity can either be viewed as an edge-based flow $f_i : E \to \mathbb{R}_+$, or as a path-based flow $f_i : \mathcal{P}_i \to \mathbb{R}_+$, where $\mathcal{P}_i$ is the set of all simple paths between s_i and t_i in G. We prefer the path-based flow since it is more convenient for treating directed and undirected graphs in a unified fashion, and also for writing the linear programs for flows and cuts in a more intuitive fashion. However, it is easier to argue polynomial-time solvability of the linear programs via edge-based flows. Given path-based flows f_i, $i = 1, \ldots, k$ for the k source-sink pairs, the total flow on an edge e is defined as $f(e) = \sum_{i=1}^{k} \sum_{p \in \mathcal{P}_i} f_i(p)$. The total flow for commodity i is $R_i = \sum_{p \in \mathcal{P}_i} f_i(p)$, where R_i is interpreted as the rate of commodity flow i. In *directed* polymatroidal networks, the flow is constrained to satisfy the following capacity constraints.

$$\sum_{e \in S} f(e) \leq \rho_v^-(S) \quad \forall v \, \forall S \subseteq \delta^-(v) \quad \text{and} \quad \sum_{e \in S} f(e) \leq \rho_v^+(S) \quad \forall v \, \forall S \subseteq \delta^+(v)$$

The constraints in *undirected* polymatroidal networks are:

$$\sum_{e \in S} f(e) \leq \rho_v(S) \quad \forall v \, \forall S \subseteq \delta(v). \tag{2.1}$$

A rate tuple $(R_1, ..., R_k)$ is said to be *achievable* if commodities $1, \ldots, k$ can be sent at rates $R_1, \ldots, R_k$ simultaneously between the corresponding source-sink pairs. For a given polymatroidal network and source-sink pairs the set of achievable rate tuples is easily seen from the above constraints to be a polyhedral set. We let $P(G, \mathcal{T})$ denote this rate region where G is the network and $\mathcal{T}$ is the set of given source-sink pairs. In the *maximum throughput multicommodity flow* problem the goal is to maximize $\sum_{i=1}^{k} R_i$ over $P(G, \mathcal{T})$. In the *maximum concurrent multicommodity flow* problem each source-sink pair has an associated demand D_i and the goal is to maximize λ such that the rate tuple $(\lambda D_1, ..., \lambda D_k)$ is achievable, that is the tuple belongs to $P(G, \mathcal{T})$. It is easy to see that both these problems can be cast as linear programming problems. The path-formulation results in an exponential (in n the number of nodes of G) number of variables and we also have an exponential number of constraints due to the polymatroid constraints at each node. However, one can use an edge-based formulation and solve the linear programs in polynomial time via the ellipsoid method and polynomial-time algorithms for submodular function minimization.

Networks with symmetric demands: In directed polymatroidal networks we are primarily interested in *symmetric demands*: node s_i intends to communicate with t_i and node t_i intends to communicate with s_i at the same *rate*. Conceptually one can reduce this to the general setting by having two commodities (s_i, t_i) and (t_i, s_i) for a pair $s_i t_i$ and adding a constraint that ensures their rates are equal. To be technically consistent with previous work we do the following. We will assume that we are given k unordered source-sink pairs $s_1 t_1, \ldots, s_k t_k$. Now consider the $2k$ ordered pairs $(s_1, t_1), \ldots, (s_k, t_k), (t_1, s_1), \ldots, (t_k, s_k)$. We are interested in achievable rate tuples of the form $(R_1, \ldots, R_k, R'_1, \ldots, R'_k)$ where $R'_i = R_i$. In the maximum throughput setting we maximize $\sum_{i=1}^{k}(R_i + R'_i)$. Note that even though the rates for (s_i, t_i) and (t_i, s_i) are the same, the flow paths along which they route can be different. In the maximum concurrent flow setting both (s_i, t_i) and (t_i, s_i) have a common demand D_i and we find the maximum λ such that rate tuple $(\lambda D_1, ..., \lambda D_k, \lambda D_1, ..., \lambda D_k)$ is achievable for the pairs $(s_1, t_1), \ldots, (s_k, t_k), (t_1, s_1), \ldots, (t_k, s_k)$.

2.2.2 Cuts

The multicommodity flow problems have natural dual cut problems associated with them. Given a graph $G = (V, E)$ and a set of edges $F \subseteq E$ we say that the ordered node pair (s, t) is separated by F if there is no path from s to t in the graph $G[E \setminus F]$. In directed graphs F may separate (s, t) but not (t, s). In undirected graphs we say that F separates the unordered node pair st if s and t are in different connected components of $G[E \setminus F]$. In the standard network model the cost of a cut defined by a set of edges F is simply $\sum_{e \in F} c(e)$ where $c(e)$ is the cost of e (capacity in the primal flow network) . In polymatroid networks the cost of F is defined in a more involved fashion. Each edge (u, v) in F is assigned to either u or v; we say that an assignment of edges to nodes $g : F \to V$ is *valid* if it satisfies this restriction. A valid assignment partitions F into sets $\{g^{-1}(v) \mid v \in V\}$ where $g^{-1}(v)$ (the pre-image of v) is the set of edges in F assigned to v by g. For a given valid assignment g of F the cost of the cut $\nu_g(F)$ is defined as

$$\nu_g(F) := \sum_v \left(\rho_v^-(\delta^-(v) \cap g^{-1}(v)) + \rho_v^+(\delta^+(v) \cap g^{-1}(v))\right).$$

In undirected graphs the cost for a given assignment is $\sum_v \rho_v(g^{-1}(v))$.

Given a set of edges F we define its cost to be the minimum over all possible valid assignments of F to nodes, the expression for the cost as above. We give a formal definition below.

Definition 1. Cost of edge cut: *Given a directed polymatroid network $G = (V, E)$ and a set of edges $F \subseteq E$, its cost denoted by $\nu(F)$ is*

$$\min_{g: F \to V,\, g \text{ valid}} \sum_v \left(\rho_v^-(\delta^-(v) \cap g^{-1}(v)) + \rho_v^+(\delta^+(v) \cap g^{-1}(v))\right). \tag{2.2}$$

In an undirected polymatroid network $\nu(F)$ is

$$\min_{g:F\to V,\ g\ \mathit{valid}} \sum_v \rho_v(g^{-1}(v)). \tag{2.3}$$

Lemma 1. *The cut cost function is sub-additive, that is, $\nu(F \cup F') \leq \nu(F) + \nu(F')$ for all $F, F' \subseteq E$.*

Although not obvious, ν can be evaluated in polynomial time via an algorithm to compute an s-t maximum flow problem in a polymatroid network. We do not, however, rely on it in this thesis.

We now define the two cuts problems of interest.

Definition 2. *Given a collection of source-sink pairs $(s_1, t_1), \ldots, (s_k, t_k)$ in $G = (V, E)$ and associated demand values $D_1, \ldots, D_k$, and a set of edges $F \subseteq E$ the demand separated by F, denoted by $D(F)$, is $\sum_{i:(s_i,t_i)\ \text{separated by}\ F} D_i$. F is a* multicut *if all the given source-sink pairs are separated by F. The* sparsity *of F is defined as $\frac{\nu(F)}{D(F)}$.*

The above definitions extend naturally to undirected graphs. Given the above definitions two natural optimization problems that arise are the following. The first is to find a multicut of minimum cost for a given collection of source-sink pairs. The second is to find a cut of minimum sparsity. These problems are NP-hard even in edge-capacitated undirected graphs and have been extensively studied from an approximation point of view [90, 48, 97, 16, 14, 2].

Lemma 2. *Given a multicommodity polymatroidal network instance, the value of the maximum throughput flow is at most the cost of a minimum multicut. The value of the maximum concurrent flow is at most the minimum sparsity.*

A key question of interest is to quantify the relative gap between the flow and cut values. These gaps are relatively well-understood in standard networks and the main aim of this thesis is to obtain results for polymatroid networks.

Networks with symmetric demands: For a directed network with symmetric demands the notion of a "cut" has to be defined appropriately. We say that a set of edges F separates a pair $s_i t_i$ if it separates (s_i, t_i) *or* (t_i, s_i). With this notion of separation, the definitions of multicut and sparsest cut extend naturally. A multicut is a set of edges F whose removal separates all the given pairs. Similarly for a set of edges F its sparsity is defined to $\nu(F)/D(F)$ where $D(F)$ is the total demand of pairs separated; note that if both (s_i, t_i) and (t_i, s_i) are separated by F we count D_i twice in $D(F)$. This is to be consistent with the definition of flows given earlier. Lemma 2 extends to the symmetric demand case with the definition of flows given for symmetric demands in the previous section.

2.3 Relaxations for Cuts

Lemma 2 gives a way to lower-bound the value of multicut and sparsest cut via corresponding flow problems. The flow problems can be cast as linear programs. The duals of these linear programs can be directly interpreted as linear programming relaxations for integer programming formulations for the cut problems. Here we take the approach of writing the formulation with a convex objective function and linear constraints; this simplifies and clarifies the constraints and aids in the analysis. For one of the cases we show the equivalence of the formulation with the dual of the corresponding flow linear program. We first discuss continuous extensions of submodular functions. We first discuss continuous extensions of submodular functions.

2.3.1 Continuous extensions of submodular functions

Given a submodular set function $\rho : 2^N \rightarrow \mathbb{R}$ on a finite ground set N, it is useful to *extend* it to a function $\rho' : [0,1]^N \rightarrow \mathbb{R}$ defined over the cube in $|N|$ dimensions. That is, we wish to assign a value for each $\mathbf{x} \in [0,1]^N$ such that $\rho'(\mathbf{1}_S) = \rho(S)$ for all $S \subseteq N$ where $\mathbf{1}_S$ is the characteristic vector of the set S. For minimizing submodular functions a natural goal is to find an extension that is convex. We describe two extensions below.

Convex closure: For a set function $\rho : 2^N \rightarrow \mathbb{R}$ (not necessarily submodular) its convex closure is a function $\tilde{\rho} : [0,1]^N \rightarrow \mathbb{R}$ with $\tilde{\rho}(\mathbf{x})$ defined as the optimum value of the following linear program:

$$\begin{aligned}
\tilde{\rho}(\mathbf{x}) &= \min \sum_{S \subseteq N} \alpha_S \rho(S) \\
\text{s.t.} & \\
\sum_{S} \alpha_S &= 1 \\
\sum_{S: i \in S} \alpha_S &= x_i \qquad \forall i \in N \\
\alpha_S &\geq 0 \qquad \forall S.
\end{aligned}$$

The function $\tilde{\rho}$ is convex for any ρ. Moreover, when ρ is submodular, for any given $\mathbf{x}$, the linear program above can be solved in polynomial time via submodular function minimization and hence $\tilde{\rho}(\mathbf{x})$ can be computed in polynomial time (assuming a value oracle for ρ). It is known and not difficult to show that if ρ is a polymatroid (monotone and $f(\emptyset) = 0$) the value of the linear program does not change if we drop the constraint that $\sum_S \alpha_S = 1$.

Lovász extension: For a set function $\rho : 2^N \to \mathbb{R}$ (not necessarily submodular) its Lovász extension [98] denoted by $\hat{\rho} : [0,1]^N \to \mathbb{R}$ is defined as follows:

$$\hat{\rho}(\mathbf{x}) = \int_0^1 \rho(\mathbf{x}^\theta) d\theta$$

, where $\mathbf{x}^\theta = \{i \mid x_i \geq \theta\}$. This is not the standard way the Lovász extension is stated but is entirely equivalent to it. The standard definition is the following. Given $\mathbf{x}$ let $i_1, \ldots, i_n$ be a permutation of $\{1, 2, \ldots, n\}$ such that $x_{i_1} \geq x_{i_2} \geq \ldots \geq x_{i_n} \geq 0$. For ease of notation define $x_0 = 1$ and $x_{n+1} = 0$. For $1 \leq j \leq n$ let $S_j = \{i_1, i_2, \ldots, i_j\}$. Then

$$\hat{\rho}(\mathbf{x}) = (1 - x_{i_1})\rho(\emptyset) + \sum_{j=1}^{n} (x_{i_j} - x_{i_{j+1}})\rho(S_j).$$

It is typical to assume that $\rho(\emptyset) = 0$ and omit the first term in the right hand side of the preceding equation. Note that it is easy to evaluate $\hat{\rho}(\mathbf{x})$ given a value oracle for ρ.

We state some well-known facts.

Lemma 3. *For a submodular set function ρ, $\tilde{\rho}(\mathbf{x}) = \hat{\rho}(\mathbf{x})$ for any $\mathbf{x} \in [0,1]^N$.*

Therefore the convex closure coincides with the Lovász extension and $\hat{\rho}(\cdot)$ is convex.

Proposition 1. *For a monotone submodular function ρ and $\mathbf{x} \leq \mathbf{x}'$ (coordinate-wise), $\hat{\rho}(\mathbf{x}) \leq \hat{\rho}(\mathbf{x}')$.*

The equivalence of $\tilde{\rho}$ and $\hat{\rho}$ also implies that an optimum solution to the linear program defining $\tilde{\rho}(\mathbf{x})$ is obtained by a solution $\bar{\alpha}$ where the support of $\bar{\alpha}$ is a chain on N (a laminar family whose tree representation is a path). In fact we have the following. Given $\mathbf{x} \in [0,1]^N$ consider the ordering of the coordinates and the associated sets as in the definition of the $\hat{\rho}(\mathbf{x})$. One can verify that $\alpha_{S_j} = x_{i_j} - x_{i_{j-1}}$ for $1 \leq j \leq n$, $\alpha_\emptyset = (1 - x_{i_n})$, and $\alpha_S = 0$ for all other sets S is an optimum solution to the linear program that defines $\tilde{\rho}(\mathbf{x})$. We will use this fact later.

2.3.2 Multicut

We now consider the multicut problem. Recall that we wish to find a subset $F \subseteq E$ such that F separates all the given source-sink pairs so as to minimize the cost $\nu(F)$. The only difference between the polymatroid networks and standard networks is in the definition of the cost. We first focus on expressing the constraint that F is a feasible set for separating the pairs. For each edge e we have a variable $\ell(e) \in [0,1]$ in the relaxation that represents whether e is cut or not. For feasibility of the cut we have the condition that for any path p from s_i to t_i (that is $p \in \mathcal{P}_i$) at least one edge in p is cut; in the relaxation this corresponds to the constraint that $\sum_{e \in p} \ell(e) \geq 1$. In other words $\text{dist}_\ell(s_i, t_i) \geq 1$ where $\text{dist}_\ell(u, v)$ is the distance between u and v with edge lengths given by $\ell(e)$ values.

We now consider the cost of the cut. Note that $\nu(F)$ is defined by valid assignments of F to the nodes, and submodular costs on the nodes. In the relaxation we model this as follows. For an edge $e = (u, v)$ we have variables $\ell(e, u)$ and $\ell(e, v)$ which decide whether e is assigned to u or v. We have a constraint $\ell(e, u) + \ell(e, v) = \ell(e)$ to model the fact that if e is cut then it has to be assigned to either u or v. Now consider a node v and the edges in $\delta^+(v)$. The variables $\ell(e, v), e \in \delta^+(v)$ in the integer case give the set of edges $S \subseteq \delta^+(v)$ that are assigned to v and in that case we can use the function $\rho_v^+(S)$ to model the cost. However, in the fractional setting the variables lie in the real interval $[0,1]$ and here we use the extension approach to obtain a convex programming relaxation;

$$
\begin{aligned}
\min \sum_v (\hat{\rho}_v^-(\mathbf{d}_v^-) + \hat{\rho}_v^+(\mathbf{d}_v^+)) \\
\ell(e,u) + \ell(e,v) = \ell(e) \qquad & e = (u,v) \in E \\
\mathrm{dist}_\ell(s_i, t_i) \geq 1 \qquad & 1 \leq i \leq k \\
\ell(e), \ell(e,u), \ell(e,v) \geq 0 \qquad & e = (u,v) \in E.
\end{aligned}
$$

$$
\begin{aligned}
\min \sum_v \hat{\rho}_v(\mathbf{d}_v) \\
\ell(e,u) + \ell(e,v) = \ell(e) \qquad & e = uv \in E \\
\mathrm{dist}_\ell(s_i, t_i) \geq 1 \qquad & 1 \leq i \leq k \\
\ell(e), \ell(e,u), \ell(e,v) \geq 0 \qquad & e = uv \in E.
\end{aligned}
$$

Figure 2.1: Lovász-extension based relaxations for multicut in directed and undirected polymatroidal networks

we can rewrite the convex program as an equivalent linear program via the definition of $\tilde{\rho}$. Let $\mathbf{d}_v^-$ be the vector consisting of the variables $\ell(e,v)$, $e \in \delta^-(v)$ and similarly $\mathbf{d}_v^+$ denote the vector of variables $\ell(e,v)$, $e \in \delta^+(v)$. The relaxation for the directed case is formally described in Fig 2.1 in the box on the left. For the symmetric demands case the relaxation is similar, but since we need to separate either (s_i, t_i) or (t_i, s_i) the constraint $\mathrm{dist}_\ell(s_i, t_i) \geq 1$ is replaced by the constraint $\mathrm{dist}_\ell(s_i, t_i) + \mathrm{dist}_\ell(t_i, s_i) \geq 1$.

For the undirected case we let $\mathbf{d}_v$ denote the vector of variables $\ell(e,v), e \in \delta(v)$ and the resulting relaxation is shown on the right in Fig 2.1.

One can replace $\hat{\rho}_v$ in the above convex programming relaxations by $\tilde{\rho}_v$ the convex closure; further, one can use the definition of $\tilde{\rho}_v$ via a linear program to convert the convex program into an equivalent linear program. The resulting linear program can be shown to be equivalent to the dual of the maximum throughput flow problem. See Section A.1 for a formal proof.

2.3.3 Sparsest cut

Now we consider the sparsest cut problem. In the sparsest cut problem we need to decide which pairs to disconnect and then ensure that we pick edges whose removal separates the chosen pairs. Moreover we are interested in the ratio of the cost of the cut to the demand separated. We follow the known formulation in the

$$
\begin{aligned}
&\min \sum_v \hat{\rho}_v^-(\mathbf{d}_v^-) + \hat{\rho}_v^+(\mathbf{d}_v^+)) \\
&\ell(e,u) + \ell(e,v) = \ell(e) \qquad e = (u,v) \in E \\
&\sum_{i=1}^k D_i \cdot \text{dist}_\ell(s_i,t_i) = 1 \\
&\ell(e), \ell(e,u), \ell(e,v) \geq 0 \qquad e = (u,v) \in E.
\end{aligned}
$$

$$
\begin{aligned}
&\min \sum_v \hat{\rho}_v(\mathbf{d}_v) \\
&\ell(e,u) + \ell(e,v) = \ell(e) \qquad e = uv \in E \\
&\sum_{i=1}^k D_i \cdot \text{dist}_\ell(s_i,t_i) = 1 \\
&\ell(e), \ell(e,u), \ell(e,v) \geq 0 \qquad e = (u,v) \in E.
\end{aligned}
$$

Figure 2.2: Relaxations for sparsest cut in directed and undirected polymatroidal networks

edge-capacitated case with the main difference, again, being in the cost of the cut. There is a variable y_i which determines whether pair i is separated or not. We again have the edge variables $\ell(e), \ell(e,u), \ell(e,v)$ to indicate whether $e = (u,v)$ is cut and whether e's cost is assigned to u or v. If pair i is to be separated to the extent of y_i we ensure that $\text{dist}_\ell(s_i,t_i) \geq y_i$. To express sparsity, which is defined as a ratio, we normalize the demand separated to be 1. Fig 2.2 has a formal description on the left for the directed case. For the symmetric demands case we have essentially the same relaxation; the constraint $\sum_i D_i \text{dist}_\ell(s_i,t_i) = 1$ is replaced by the constraint $\sum_i D_i(\text{dist}_\ell(s_i,t_i) + \text{dist}_\ell(t_i,s_i)) = 1$.

The relaxation for the undirected case is shown on the right in Fig 2.2 where $\mathbf{d}_v$ is the vector of variables $\ell(e,v), e \in \delta(v)$.

2.4 Flow-Cut Gaps in Directed Polymatroidal Networks

In this section we consider flow-cut gaps in directed polymatroidal networks. We show via a reduction that these gaps can be related to corresponding gaps in directed edge-capacitated networks that have been well-studied. We note that this reduction is specific to directed graphs and does not apply to undirected polymatroidal networks. The embedding based approach for the undirected case that we discuss in Section 2.5 is also applicable to directed graphs.

The reduction is similar at a high level for both gap questions of interest and is based on the relaxations for the two cut problems that we described in Section 2.3. We take a feasible fractional solution for relaxation of the cut problem in question and produce an instance of a cut problem in an edge-capacitated network and a feasible fractional solution to the corresponding cut problem. We also provide a correspondence between feasible integer solutions to the edge-capacitated network instance and the original problem such that the cost of the solution is preserved. These correspondences allow us to translate known gap results for the edge-capacitated networks to polymatroidal networks.

2.4.1 Details of the reduction

Let $G = (V, E)$ be a directed graph and let $\ell : E \to \mathbb{R}_+$ be a length function on the edges. We let $\text{dist}_\ell(u, v)$ be the shortest path distance from u to v in G with edge lengths ℓ. Moreover, for each edge (u, v) let $\ell(e, u)$ and $\ell(e, v)$ be two non-negative numbers such that $\ell(e) = \ell(e, u) + \ell(e, v)$. For a node v let $\mathbf{d}_v^+$ be the vector of $\ell(e, v)$ values for all edges $e \in \delta^+(v)$ and similarly $\mathbf{d}_v^-$ is the vector of $\ell(e, v)$ values for edges in $\delta^-(v)$. In the polymatroidal setting the cost induced by the edge length variables is given by $\sum_{v \in V}(\hat{\rho}^-(\mathbf{d}_v^-) + \hat{\rho}^+(\mathbf{d}_v^+))$. Note that for multicut we have that $\text{dist}_\ell(s_i, t_i) \geq 1$ for each demand pair (s_i, t_i) while in sparsest cut we are interested in the ratio of the cost to $\sum_i D_i \cdot \text{dist}_\ell(s_i, t_i)$. We now describe the construction of a graph $H = (V_H, E_H)$ where $V_H = V \uplus V'$ (that is the nodes of G are also in H) and an edge length function $\ell' : E_H \to \mathbb{R}_+$ such that $\text{dist}_\ell(u, v) = \text{dist}_{\ell'}(u, v)$ for all $u, v \in V$; that is the distances between nodes in V are the same in G and H. We also create an edge-cost (or capacity in the primal sense) function $c : E_H \to \mathbb{R}_+$. The construction will also establish the correspondence of cuts in G and H and their costs.

The graph $H = (V \uplus V', E_H)$ is constructed as follows. To aid the reader we first describe the idea of the construction at a high-level. Consider a node $v \in V$ and the in-coming edges $\delta^-(v)$ and out-going edges $\delta^+(v)$. In H we have nodes of V and build an in-tree T_v^- and an out-tree T_v^+ that are rooted at v. The leaves

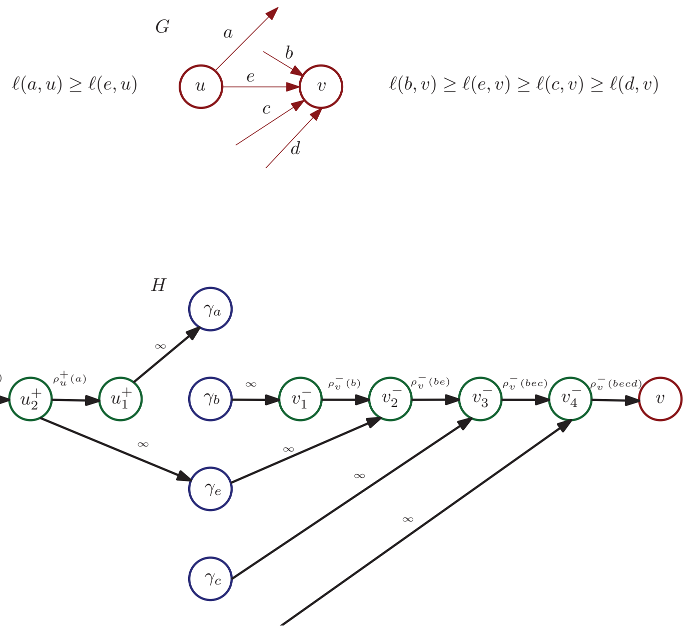


Figure 2.3: Illustration of the reduction. Only $\delta_G^+(u)$ and $\delta_G^-(v)$ are shown. The costs on edges in H are shown but not their lengths. The lengths of the infinite cost edges is 0 and $\ell'(u_2^+, u_1^+) = \ell(a,u) - \ell(e,u)$ and $\ell'(v_3^-, v_4^-) = \ell(c,v) - \ell(d,v)$.

of T_v^- are the edges in $\delta^-(v)$ the leaves of T_v^+ are the edges in $\delta^+(v)$. Note that an edge (u,v) will thus participate in T_u^+ and T_v^-. Now for the formal details. The nodes of H, denoted by V_H, consist of the nodes V of G and additional nodes V'. V' has two types of nodes. First, for each edge $e \in E$ there is a node γ_e. Second, for each node $v \in V$ we create two sets of nodes $N^-(v)$ and $N^+(v)$ where $|N^-(v)| = n_v^- = |\delta_G^-(v)|$ and $|N^+(v)| = n_v^+ = |\delta_G^+(v)|$; thus one node for each edge in $\delta^-(v) \cup \delta^-(v)$; these will be the internal nodes of the trees T_v^- and T_v^+ respectively. For notational convenience we refer to the j'th node in $N^-(v)$ as v_j^- and similarly v_j^+ for the j'th node in $N^+(v)$.

Now we describe the edge set E_H of the graph H, the edge length function $\ell' : E_H \to \mathbb{R}_+$, and the cost function $c : E_H \to \mathbb{R}_+$. The edge set is essentially prescribed by specifying the trees T_v^- and T_v^+ for each $v \in V$. Consider the vector $\mathbf{d}^-(v)$ of values $\ell(e,v)$ for $e \in \delta_G^-(v)$. Recall the definition of the Lovász extension $\hat{\rho}^-(\mathbf{d}_v^-)$. We order the edges in $\delta^-(v)$ as $e_1, e_2, \ldots, e_{n_v^-}$ where $\ell(e_j,v) \geq \ell(e_{j+1},v)$ for $1 \leq j < n_v^-$ and then $\hat{\rho}^-(\mathbf{d}_v^-) = \sum_j (\ell(e_j,v) - \ell(e_{j+1},v))\rho_v^-(S_j)$ where $S_j = \{e_1, \ldots, e_j\}$. We associate the node v_j^- with the set S_j. The edge set of T_v^- is defined as follows. For ease of notation we let $v_{n_v^-+1}^-$ represent the node v. We create a directed path $v_1^- \to v_2^- \to \ldots \to v_{n_v^-}^- \to v_{n_v^-+1}^- = v$ with edge lengths $\ell'(v_1^-, v_2^-) = \ell(e_1,v) - \ell(e_2,v), \ell'(v_2^-, v_3^-) = \ell(e_2,v) - \ell(e_3,v), \ldots, \ell'(v_{n_v^-}^-, v) = \ell(e_{n_v^-}, v) - 0$. The costs of these edges are defined as follows: $c(v_j^-, v_{j+1}^-) = \rho_v^-(S_j)$ for $1 \leq j \leq n_v^-$. For each j we add the edge (γ_{e_j}, v_j^-) with length 0 and cost ∞ (for computational purpose a sufficiently large number M would do); this connects the node γ_{e_j} corresponding to the edge e_j to v_j^- that corresponds to S_j. See Fig 2.3.

The construction of T_v^+ is quite similar except that the edge directions are reversed; assuming that the edges in $\delta^+(v)$ are ordered such that $\ell(e_1,v) \geq \ell(e_2,v) \geq \ldots \geq \ell(e_{n_v^+},v)$, we create a path $v \to v_{n_v^+}^+ \to \ldots v_2^+ \to v_1^+$ with edge lengths $\ell(e_{n_v^+},v) - 0, \ldots, \ell(e_j,v) - \ell(e_{j+1},v), \ldots, \ell(e_1,v) - \ell(e_2,v)$. The costs for the edges in this path are set to $\rho_v^+(S_{n_v^+}), \ldots, \rho_v^+(S_1)$ where $S_j = \{e_1, \ldots, e_j\}$. For each j we add an edge (v_j^+, γ_{e_j}) with length 0 and cost ∞. This finishes the description of H. We now describe various properties of the graph H. Several of these properties are straightforward from the description of the construction and we omit proofs of the easy claims.

The proposition below asserts the cost of the fractional solution in the edge-capacitated network H is the same as the cost of the fractional solution in the polymatroidal network G.

Proposition 2. $\sum_{e \in E_H} c(e) \cdot \ell'(e) = \sum_{v \in V} (\hat{\rho}^-(\mathbf{d}_v^-) + \hat{\rho}^+(\mathbf{d}_v^+))$.

Proposition 3. *For any edge $e \in \delta_G^-(v)$ the length of the unique path in T_v^- from the node γ_e to v is equal to $\ell(e, v)$. Similarly for $e \in \delta_G^+(v)$, the length of the unique path in T_v^+ from the node v to the node γ_e is equal to $\ell(e, v)$.*

We now establish a correspondence between paths in G and H that connect nodes in V. Let $e = (u, v)$ be an edge in G. We obtain a canonical path $q(u, v)$ from u to v in H as follows: concatenate the unique path from u to γ_e in T_v^+ with the unique path from γ_e to v in T_v^-. For any two nodes $s, t \in V$ let $\mathcal{P}_G(s, t)$ be the set of (simple) s-t paths on G and similarly $\mathcal{P}_H(s, t)$ be the paths in H. We create a map $g : \mathcal{P}_G(s, t) \to \mathcal{P}_H(s, t)$ as follows. Consider a path $p \in \mathcal{P}_G(s, t)$; we obtain a path $p' \in \mathcal{P}_H(s, t)$ corresponding to p as follows. We replace each edge $(u, v) \in p$ by the canonical path $q(u, v)$.

Lemma 4. *The map g is a bijection. Moreover, for any two nodes $u, v \in V$, $dist_{\ell'}(u, v) = dist_{\ell}(u, v)$.*

Now we establish a correspondence between cuts in G and H. For a given set of edges $F \subseteq E$ let $\text{sep}_G(F)$ be set of node pairs in $V \times V$ separated by F in the graph G. Similarly for a set of edges $F' \subseteq E_H$ let $\text{sep}_H(F')$ be the set of node pairs in $V \times V$ separated by F' in the graph H. We say that a set of edges F is minimal with respect to separating node pairs if there is no proper subset of F that separates the same node pairs as F.

Proposition 4. *Let $F' \subseteq E_H$ be minimal with respect to separating node pairs in $V \times V$ and of finite cost. Then for any $v \in V$, F' contains at most one edge from T_v^- and at most one edge from T_v^+.*

Proof. Consider a node v and edge-sets $F' \cap T_v^-$ and $F' \cap T_v^+$. For an edge $e \in E$ there is a node $\gamma_e \in V_H$ and there is exactly one edge coming into γ_e and exactly one edge going out of γ_e and both are of infinite cost. Therefore, if F' is of finite cost, $F' \cap T_v^-$ consists of some edges in the path $v_1^- \to v_2^- \ldots \to v_{n_v^-}^- \to v$ contained in T_v^-. Since the only way to reach v is through T_v^- it follows that if F' contains an edge (v_j^-, v_{j+1}^-) then it is redundant to remove an edge (v_i^-, v_{i+1}^-) for $i < j$. Thus minimality of F' implies F' contains exactly one edge from T_v^-. The reasoning for T_v^+ is similar. $\square$

Lemma 5. *Let $F' \subseteq E_H$ be minimal with respect to separating node pairs in $V \times V$ and of finite cost. There exists a set of edges $F \subseteq E$ such that $sep_G(F) \supseteq sep_H(F')$ and $\nu(F) \leq c(F')$.*

Proof. Given a minimal F' we obtain a set of edges $F \subseteq E$ as follows. From the proof of Proposition 4 we see that for any node v, F' contains at most one edge from T_v^- and in particular if it contains an edge then it is an edge (v_j^-, v_{j+1}^-) for some $1 \leq j \leq n_v^-$ (for simplicity we identify v with $v_{n_v^-+1}^-$). Suppose there is such an edge $e' = (v_j^-, v_{j+1}^-)$ in F'. Note that e' corresponds to the set $S_j = \{e_1, \ldots, e_j\}$ of edges in $\delta_G^-(v)$ ordered in increasing order by $\ell(e, v)$ values. We add S_j to F and assign these edges to v in upper bounding $\nu(F)$: by construction $c(e') = \rho_v^-(S_j)$. We do a similar procedure if $e' \in F \cap T_v^+$. It follows that the edge set F that we construct satisfies the property that $\nu(F) \leq c(F')$.

We now show that $\text{sep}_G(F) \supseteq \text{sep}_H(F')$. Consider a pair (s, t) such that s is separated from t by F' in H. Suppose (s, t) is not separated by F in G. Let p be an s-t path that remains in $G \setminus F$. From Proposition 3 there is a unique path $g(p) \in \mathcal{P}_H(s, t)$. For every edge $e = (u, v) \in p$ consider the canonical path $q(u, v)$ in H. Since e is not in F it implies that u can reach γ_e in $H \setminus F'$ and that γ_e can reach v in $H \setminus F'$. This means that $q(u, v)$ exists in $H \setminus F'$. This would imply that $g(p)$ exists in $H \setminus F'$ contradicting that assumption that (s, t) is separated by F'. $\square$

We summarize the properties of the reduction. We assume that we have a polymatroidal network $G = (V, E)$ with k demand pairs $(s_i, t_i), \ldots, (s_k, t_k)$ with associated demand values $D_1, \ldots, D_k$. For all the cut problems of interest, the relaxations in Section 2.3 produce a length function $\ell : E \to \mathbb{R}_+$ and for each $e = (u, v)$ associated non-negative values $\ell(e, u)$ and $\ell(e, v)$ such that $\ell(e) = \ell(e, u) + \ell(e, v)$. As before we use $\mathbf{d}_v^-$ and $\mathbf{d}_v^+$ to denote the vector of $\ell(e, v)$ values for the incoming and outgoing edges at v. The reduction produces an edge-capacitated network $H = (V_H, E_H)$ with the following properties:

- each node of V is a node in V_H
- for all $u, v \in V$, $\text{dist}_\ell(u, v) = \text{dist}_{\ell'}(u, v)$
- $\sum_{e \in E_H} c(e)\ell'(e) = \sum_{v \in V} (\hat{\rho}_v^-(\mathbf{d}_v^-) + \hat{\rho}_v^+(\mathbf{d}_v^+))$
- for any set of edges $F' \subseteq E_H$ there is a corresponding set $F \subseteq E$ such that $\text{sep}_G(F) \supseteq \text{sep}_H(F')$ and $\nu(F) \leq c(F')$.

We also note that the reduction can be carried out in polynomial time. Moreover, given a set $F' \subseteq E_H$ a set $F \subseteq E$ that satisfies the last property in the list above can be found in polynomial time.

We build on the reduction to obtain flow-cut gap results, all of which are based on using the relaxations from Section 2.3 which are dual to the corresponding flow problems. We argue via the reduction and known results on edge-capacitated networks that there exist integral cuts within some factor α of the fractional solution.

2.4.2 Multicut

We consider the multicut problem for arbitrary demand pairs as well as symmetric demands. The relaxation satisfies the constraint that $\text{dist}_\ell(s_i, t_i) \geq 1$ for each demand pair (s_i, t_i). The reduction from the preceding section produces a graph $H = (V_H, E_H)$ and a fractional solution $\ell' : E_H \to \mathbb{R}_+$ such that $\text{dist}_{\ell'}(s_i, t_i) \geq 1$. We note that ℓ' is a feasible solution for the standard distance based relaxation for multicut in edge-capacitated networks which is the dual for the maximum throughput multicommodity flow problem. The integrality gap of this relaxation has been studied and several results are known. Let $\beta = \sum_{e \in E_H} c(e)\ell'(e)$ be the fractional solution value. Then one can obtain an integral multicut F' with cost $c(F')$ that can be bounded in terms of β. We summarize the known results.

- Cheriyan, Karloff and Rabani [26] showed that there exists an F' such that $c(F') \leq O(1) \cdot \beta^3$; this was improved by Gupta [54] to show the existence of a multicut F' such that $c(F') \leq O(1) \cdot \beta^2$. These results hold under the assumption that $c(e) \geq 1$ for all e.

- Agrawal, Alon and Charikar [2] improving the results in [26, 54] showed the existence of a cut F' such that $c(F') = \tilde{O}(n^{11/23}) \cdot \beta$. Here n is the number of nodes in the graph.

- Saks, Samorodnitsky and Zosin [130] showed that there exist instances on which every integral multicut has a value $\Omega(k) \cdot \beta$.

- Chuzhoy and Khanna [28] showed that there exist instances on which every multicut has a value $\tilde{\Omega}(n^{1/7}) \cdot \beta$. Further, they showed that the multicut problem is hard to approximate to within a factor of $\Omega(2^{\log^{1-\epsilon} n})$ unless $NP \subseteq ZPP$.

Since polymatroidal networks generalize edge-capacitated networks it follows that all the lower bounds in the above hold for the polymatroidal network case as well. The reduction also allows us to obtain an upper-bound for polymatroidal networks. We have to be careful when using bounds that depend on the number of nodes in the graph. The reduction takes G with n nodes and m edges and produces an edge-capacitated graph H with $n + 2m$ nodes. In the worst case H has $\Omega(n^2)$ nodes. We thus obtain the following theorem.

Theorem 1. *In a directed polymatroidal network G on n nodes, for any given multicommodity flow instance with k pairs, if β is the maximum throughput multicommodity flow then:*

- *There is a feasible multicut F' such that $\nu(F') \leq O(1) \cdot \beta^2$ assuming that ρ_v^+ and ρ_v^- are integer valued for all $v \in V$.*
- *There is a feasible multicut F' such that $\nu(F') \leq \tilde{O}(n^{22/23}) \cdot \beta$.*

Moreover, there exist polynomial-time algorithms to find multicuts guaranteed as above.

Symmetric demands: We now consider the symmetric demand case when a multicut corresponds to separating (s_i, t_i) or (t_i, s_i) for a given demand pair $s_i t_i$. The relaxation for this has a constraint that $\text{dist}_\ell(s_i, t_i) + \text{dist}_\ell(t_i, s_i) \geq 1$. In contrast to the strong negative results for the general multicut problem, polylogarithmic upper bounds on flow-cut gaps are known for symmetric demands in standard networks. In particular Klein et al. [77] show that if β is the cost of a fractional solution then there exists an integral multicut of cost $O(\log^2 k) \cdot \beta$. Even et al. [41] showed the existence of a multicut of cost $O(\log n \log\log n) \cdot \beta$. Note that these bounds are incomparable in that depending on the relationship between k and n one is better than the other. It is also known that there exist instances on which the gap is at least $\Omega(\log n)$. Via the reduction we obtain the following.

Theorem 2. *In a directed polymatroidal network G on n nodes, for any given multicommodity flow instance with symmetric demands on k pairs, the minimum multicut is $O(\min\{\log^2 k, \log n \log\log n\}) \cdot \beta$ where β is maximum throughput multicommodity flow for the symmetric demands.*

Remark 1. *The flow-cut gap in polymatroidal networks for multiterminal flows*[3] *can be shown to be 2 via the reduction and the result of Naor and Zosin [109].*

[3] In multiterminal flows we have a set of k terminals $\{s_1, s_2, \ldots, s_k\}$ and flow can be sent between any pair of terminals; the goal is to maximize the total flow. The corresponding cut is referred to as multiterminal cut or multiway cut in which the goal is to remove a minimum-cost set of edges to disconnect every (ordered) pair of terminals.

2.4.3 Sparsest cut

Now we consider the sparsest cut problem where the goal is to find a set of edges F to minimize $\nu(F)/D(F)$ where $D(F)$ is the total demand of the pairs separated by F. The relaxation corresponds to finding edge length variables ℓ to minimize the fractional cost subject to the constraint that $\sum_i D_i \cdot \text{dist}_\ell(s_i, t_i) = 1$. Via the reduction we produce an edge-capacitated network H such that $\sum_i D_i \cdot \text{dist}_{\ell'}(s_i, t_i) = 1$ and with the fractional cost preserved. In edge-capacitated networks there is a generic strategy that translates the flow-cut gap for multicut into a flow-cut gap for sparsest cut at an additional loss of an $O(\log \sum_i D_i)$ factor due to Kahale [65] (see also [136]); this has been refined via a more intricate analysis in [122] to lose only an $O(\log k)$ factor although one needs to apply it carefully. In [2] a simple reduction that loses an $O(\log n)$ factor is given (this builds on [65]). For directed graphs the known-gaps for sparsest cut are essentially based on using the corresponding gap for multicut and translating via the above mentioned schemes. We thus obtain the following results.

Theorem 3. *In a directed polymatroidal network G on n nodes, for any given multicommodity flow instance with k pairs, if β is the value of the maximum concurrent flow then there is a cut of sparsity at most $\tilde{O}(n^{22/23}) \cdot \beta$.*

Theorem 4. *In a directed polymatroidal network G on n nodes, for any given multicommodity flow instance with symmetric demands on k pairs, there is a cut of sparsity $O(\min\{\log^3 k, \log^2 n \log\log n\}) \cdot \beta$ where β is maximum concurrent flow.*

2.5 Flow-Cut Gaps in Undirected Polymatroidal Networks

In this section we consider flow-cut gaps in undirected polymatroidal networks. As we already noted, node-capacitated flows are a special case of polymatroidal flows. We show that line embeddings with low average distortion introduced by Matousek and Rabinovich [104] (and further studied in [123]) are useful for bounding the gap between the maximum concurrent flow and sparsest cut; we are inspired to make this connection from [45] who considered node-capacitated flows. For multicut we show that the region growing technique from [90] that was used in [48] for edge-capacitated multicut can be adapted to the polymatroidal setting. These techniques are also applicable to directed graphs — we defer a more detailed discussion.

Furthermore, for sparsest cut in undirected graphs with product demands, we show that a relaxation based on semi-definite programming gives a tighter $O(\sqrt{\log k})$ approximation algorithm. We prove this using low-average-distortion ($O(\sqrt{\log k})$) embeddings of negative-type metrics into the line (as shown in [14] [15]).

2.5.1 Maximum concurrent flow and sparsest cut

We start with the definition of line embeddings and average distortion.

Let (V, d) be a finite metric space. A map $g : V \to \mathbb{R}$ is an embedding of V into a line; it is a *contraction* (also called 1-Lipschitz) if for all $u, v \in V$,

$$|g(u) - g(v)| \leq d(u, v).$$

Given a demand function $w : V \times V \to \mathbb{R}_+$ and a contraction $g : V \to \mathbb{R}$, its *average distortion* with respect to w is defined as

$$\text{avgd}_w(g) = \frac{\sum_{u,v \in V} w(u, v) \cdot d(u, v)}{\sum_{u,v \in V} w(u, v) \cdot |g(u) - g(v)|}$$

The following theorem is implicit in [19]; see [45] for a sketch.

Theorem 5 (Bourgain [19])**.** *For every n-point metric space (V, d) and every weight function $w : V \times V \to \mathbb{R}_+$ there is a polynomial-time computable contraction $g : V \to \mathbb{R}$ such that avgd$_w(g) = O(\log n)$. Moreover, if the support of w is k there is a map g such that avgd$_w(g) = O(\log k)$.*

Using the above we prove the following.

Theorem 6. *In undirected polymatroidal networks, for any given multicommodity flow instance with k pairs, the ratio between the value of the sparsest cut and the value of the maximum concurrent flow is $O(\log k)$. Moreover, there is an efficient algorithm to compute an $O(\log k)$ approximation to the sparsest cut problem.*

Recall the relaxation for the sparsest cut from Section 2.3.3 and the associated notation. To prove the theorem we consider an optimum solution to the relaxation and show the existence of a cut whose sparsity is $O(\log k)$ times the value of the relaxation. Let (V, d) be the metric induced on V by shortest path distances in the graph with edge lengths given by $\ell : E \to \mathbb{R}_+$ from the optimum fractional solution. Let $g : V \to \mathbb{R}$ be line embedding guaranteed by Theorem 5 with respect to d and the weight function given by the demands D_i; that is $w(s_i, t_i) = D_i$ for a demand pair and is 0 for any pair of nodes that do not correspond to a demand. Without loss of generality we can assume that g maps V to the interval $[0, \beta]$ for some $\beta > 0$. For $\theta \in (0, \beta)$ let $S_\theta = \{u \mid g(u) \leq \theta\}$. We show that there is a θ such that $\delta(S_\theta)$ is an approximately good sparse cut. Let $D(\delta(S_\theta))$ be the total demand of pairs separated by S_θ, that is $D(\delta(S_\theta)) = \sum_{i:S_\theta \text{ separates } s_i t_i} D_i$.

Lemma 6.
$$\int_0^\beta D(\delta(S_\theta)) d\theta = \Omega\left(\frac{1}{\log k}\right).$$

Proof. From the definition of $D(\delta(S_\theta))$,

$$\begin{aligned} \int_0^\beta D(\delta(S_\theta)) d\theta &= \int_0^\beta \Big(\sum_{i:S_\theta \text{ separates } s_i t_i} D_i \Big) d\theta \\ &= \sum_{i=1}^k D_i \cdot \int_0^\beta \mathbf{1}_{S_\theta \text{ separates } s_i t_i} d\theta = \sum_{i=1}^k D_i \cdot |g(s_i) - g(t_i)|. \end{aligned} \tag{2.4}$$

From the properties of g,

$$\frac{\sum_i D_i \cdot d(s_i, t_i)}{\sum_i D_i \cdot |g(s_i) - g(t_i)|} \leq O(\log k).$$

We have the constraint $\sum_i D_i \cdot d(s_i, t_i) = 1$ from the LP relaxation; this combined with the above inequality proves the lemma. □

The main insight in the proof is the following lemma. A version of the lemma also holds for directed graphs that we address in a remark following the proof.

Lemma 7.
$$\int_0^\beta \nu(\delta(S_\theta)) d\theta \leq 2 \sum_u \hat{\rho}_u(\mathbf{d}_u).$$

Proof. Consider an edge $uv \in \delta(S_\theta)$ and for simplicity assume $g(u) < g(v)$. The length of e in the embedding is $\ell'(e) = |g(v) - g(u)| \leq \ell(e)$. The edge $(u, v) \in \delta(S_\theta)$ iff θ is in the interval $[g(u), g(v)]$. Note that the cost $\nu(\delta(S_\theta))$ is in general a complicated function to evaluate. We upper bound $\nu(\delta(S_\theta))$ by giving an explicit way to assign $e = uv$ to either u or v as follows. Recall that in the relaxation

$\ell(e) = \ell(e,u) + \ell(e,v)$ where $\ell(e,u)$ and $\ell(e,v)$ are the contributions of u and v to e. Let $r = \frac{\ell(e,u)}{\ell(e)}$ and let $\ell'(e,u) = r\ell'(e)$ and $\ell'(e,v) = (1-r)\ell'(e)$. We partition the interval $[g(u), g(v)]$ into $[g(u), g(u) + \ell'(e,u))$ and $[g(u) + \ell'(e,u), g(v)]$; if θ lies in the former interval we assign e to u, otherwise we assign e to v. This assignment procedures describes a way to upper bound $\nu(\delta(S_\theta))$ for each θ. Now we consider the quantity $\int_0^\beta \nu(\delta(S_\theta))d\theta$ and upper bound it as follows.

Consider a node u and let $L_u = \{uv \in \delta(u) \mid g(v) < g(u)\}$ be the set of edges uv that go from u to the left of u in the embedding g. Similarly $R_u = \{uv \in \delta(u) \mid g(v) \geq g(u)\}$. Note that L_u and R_u partition $\delta(u)$. Let $\mathbf{d}'_u$ be the vector of dimension $|\delta(u)|$ consisting of the values $\ell'(e,u)$ for $e \in \delta(u)$. We obtain $\mathbf{d}^L_u$ from $\mathbf{d}'_u$ by setting the values for $e \in R_u$ to 0 and similarly $\mathbf{d}^R_u$ from $\mathbf{d}'_u$ by setting the values for $e \in L_u$ to 0. Since $0 \leq \ell'(e,u) \leq \ell(e,u)$ for each $e \in \delta(u)$ we see that $\mathbf{d}'_u \leq \mathbf{d}_u$ and (component wise) and hence $\mathbf{d}^L_u \leq \mathbf{d}_u$ and $\mathbf{d}^R_u \leq \mathbf{d}_u$. Since ρ_u is monotone we have that $\hat{\rho}_u(\mathbf{d}^L_u) \leq \hat{\rho}_u(\mathbf{d}_u)$ and $\hat{\rho}_u(\mathbf{d}^R_u) \leq \hat{\rho}_u(\mathbf{d}_u)$ (see Proposition 1).

We claim that

$$\int_0^\beta \nu(\delta(S_\theta))d\theta \leq \sum_{u \in V} (\hat{\rho}_u(\mathbf{d}^L_u) + \hat{\rho}_u(\mathbf{d}^R_u)),$$

which would prove the lemma.

To see the claim consider some fixed θ and $\nu(\delta(S_\theta))$. Fix a node u and consider the edges in $\delta(u) \cap S_\theta$ assigned to u by the procedure we described above; call this set $A_{\theta,u}$. First assume that $\theta < g(u)$. Then the edges assigned to u by the procedure, denoted by $A_{\theta,u} = \{e \in L_u \mid \theta > g(u) - \ell'(e,u)\}$. Similarly, if $\theta > g(u)$, $A_{\theta,u} = \{e \in L_u \mid \theta < g(u) + \ell'(e,u)\}$. From these definitions we have

$$\int_0^\beta \nu(\delta(S_\theta))d\theta \leq \sum_{u \in V} \int_0^\beta \rho_u(A_{\theta,u})d\theta.$$

For a fixed node u,

$$\int_0^\beta \rho_u(A_{\theta,u})d\theta \quad = \quad \int_0^{g(u)} \rho_u(A_{\theta,u})d\theta + \int_{g(u)}^\beta \rho_u(A_{\theta,u})d\theta.$$

Let $L_u = \{e_1, e_2, \ldots, e_h\}$ where $0 \leq \ell'(e_1, u) \leq \ell'(e_2, u) \leq \ldots \leq \ell'(e_h, u)$. Then

$$\int_0^{g(u)} \rho_u(A_{\theta,u})d\theta = \sum_{j=1}^{h}(\ell'(e_j, u) - \ell'(e_{j-1}, u))\rho(\{e_1, e_2, \ldots, e_j\}).$$

The right-hand side of the above, is, by construction and the definition of the Lovász extension, equal to $\hat{\rho}_u(\mathbf{d}_u^L)$. Similarly, $\int_{g(u)}^{\beta} \rho_u(A_{\theta,u})d\theta = \hat{\rho}_u(\mathbf{d}_u^R)$. $\square$

Remark 2. *An examination of the proof of the above lemma explains the factor of* 2 *on the right hand side; the edges in* $\delta(v)$ *can be both to the left and right of* v *in the line embedding and each side contributes* $\hat{\rho}_u(\mathbf{d}_v)$ *to the cost. This is related to the technical issue about undirected polymatroid networks where the flow through* v *takes up capacity on two edges incident to* v*. For directed graphs one can prove a statement of the form below where* $\delta^+(S_\theta)$ *is set of edges leaving* S_θ*. Notice that there is no factor of* 2 *since one treats the incoming and outgoing edges separately.*

$$\int_0^{\beta} \nu(\delta^+(S_\theta))d\theta \leq \sum_u (\hat{\rho}_u^-(\mathbf{d}_u^-) + \hat{\rho}_u^+(\mathbf{d}_u^+)).$$

The above statement gives an embedding proof of the maxflow-mincut theorem for single-commodity directed polymatroidal networks and has other applications.

We now finish the proof of Theorem 6 via the preceding two lemmas.

$$\begin{aligned}
\min_{\theta \in (0,\beta)} \frac{\nu(\delta(S_\theta))}{D(\delta(S_\theta))} &\leq \frac{\int_0^{\beta} \nu(\delta(S_\theta))d\theta}{\int_0^{\beta} D(\delta(S_\theta))d\theta} \\
&\leq 2\sum_u \hat{\rho}_u(\mathbf{d}_u) \cdot O(\log k) = O(\log k)\sum_u \hat{\rho}_u(\mathbf{d}_u).
\end{aligned}$$

The above shows that the sparsity of S_θ for some θ is at most $O(\log k)$ times $\sum_u \hat{\rho}_u(\mathbf{d}_u)$ which is the value of the relaxation. Given a line embedding g there are only $n-1$ distinct cuts of interest and one can try all of them to find the one with the smallest sparsity. The efficiency of the algorithm therefore depends on complexity of the solving the fractional relaxation and the complexity of finding a line embedding guaranteed by Theorem 5. Since both have polynomial time algorithms, one can find an $O(\log k)$ approximation to the sparsest cut in polynomial time.

Remark 3. *Node-weighted flows and cuts/separators can be cast as special cases of flows and cuts in polymatroid networks. Our algorithm produces edge-cuts from line embeddings in a simple way even for node-weighted problems — the ν cost of the edge-cut automatically translates into an appropriate node-weighted cut. In contrast, the algorithm in [45] has to solve several instances of s-t separator problems in auxiliary graphs obtained from the line embedding.*

Sparsest bi-partition cut

Till now, we worked with general edge cuts, but for certain applications, it is necessary to work with a special type of edge cut called the bi-partition cut. In an undirected polymatroidal network, an edge-cut F is said to be a *bi-partition cut* if there exists a set $S \subseteq V$ such that $F := \{e = uv : u \in S, v \in S^c$ or $v \in S, u \in S^c\}$; we denote such an edge cut by F_S. In the case of edge-capacitated undirected networks, it is well known that for finding the sparsest cut, it is sufficient to restrict the edge cuts to bi-partition cuts. While this no longer continues to be true for polymatroidal networks, a factor 2 gap can indeed be shown between the sparsest cut and the sparsest cut restricted to only bi-partition cuts.

Theorem 7. *Given any edge cut for an undirected polymatroidal network, there exists a bi-partition cut whose sparsity is at most* 2 *times the sparsity of the edge cut. Furthermore this factor is tight.*

Proof. The proof is deferred to Section A.3. □

Now, Theorem 6 and Theorem 7 together imply a logarithmic gap between maximum concurrent flow and sparsest bi-partition cut. This is formally stated in the following corollary.

Corollary 1. *In undirected polymatroidal networks, for any given multicommodity flow instance with k pairs, the ratio between the value of the* sparsest bi-partition cut *and the value of the maximum concurrent flow is $O(\log k)$.*

2.5.2 Maximum throughput flow and multicut

We prove the following theorem in this section.

Theorem 8. *In undirected polymatroidal networks, for any given multicommodity flow instance with k pairs, the ratio between the value of the minimum multicut* *and the value of the maximum throughput flow is $O(\log k)$. Moreover, there is an efficient algorithm to compute an $O(\log k)$ approximation to the minimum multicut problem.*

We recall the relaxation for the minimum multicut problem from Section 2.3.2. Consider an optimum solution to the relaxation given by edge lengths $\ell(e), e \in E$ and the partition of $\ell(e)$ for each $e = uv$ between u and v given by the variables $\ell(e, u)$ and $\ell(e, v)$. We will show that there exists a multicut $F \subseteq E$ for the given pairs such that $\nu(F) = O(\log k)(\sum_v \hat{\rho}_v(\mathbf{d}_v))$.

By slightly generalizing the proof of Lemma 7 we obtain the following.

Lemma 8. *Let $g : V \to [0, \beta]$ be a contraction, let $0 \le a_0 \le a < b \le b_0 \le \beta$ and $S_\theta = \{u \mid g(u) < \theta\}$. Suppose for every edge $e = uv \in \cup_{\theta \in [a,b]} \delta(S_\theta)$, $g(u)$ and $g(v)$ are both in $[a_0, b_0]$. Then,*

$$\int_a^b \nu(\delta(S_\theta))d\theta \le 2 \sum_{v:g(v)\in[a_0,b_0]} \hat{\rho}_v(\mathbf{d}_v).$$

Proof. The proof is very similar to the proof of Lemma 7, except that to upper bound the left-hand side in the statement of the lemma, we only need to consider edges that are in the set $\cup_{\theta\in[a,b]}\delta(S_\theta)$. The condition in the lemma assures us that any node that is involved in $\delta(S_\theta)$ has to lie within the interval $[a_0, b_0]$. Thus, it is sufficient to consider the set of nodes $v : g(v) \in [a_0, b_0]$ in the integral on the right hand side. The proof is written out in detail in Sec. A.2. $\square$

Given a graph G with edge lengths $\ell : E \to \mathbb{R}_+$, a node v and radius r, let $B^\ell_G(v, r) = \{u \mid \text{dist}_\ell(v, u) \le r\}$ denote the ball of radius r around v according to edge lengths ℓ. We omit ℓ and G if they are clear from the context. For a set of nodes $X \subseteq V$ we let $\text{vol}(X) = \sum_{v\in X} \hat{\rho}_v(\mathbf{d}_v)$ denote the total contribution of the nodes in X to the objective function.

Lemma 9. *Let $\delta < 1$ and suppose $\ell(e) < \frac{\delta}{2\log k}$ for all e. Then, for any given node s and $k \ge 2$ there exists a $r \in [0, \delta)$ such that $\nu(\delta(B(s, r)) \le a \log k \cdot \frac{1}{\delta}(\text{vol}(B(s, r)) + \text{vol}(V)/k)$, with $a = 28$.*

Proof. For simplicity we assume here that $\log k$ is an integer multiple of 3. Order the nodes in increasing order of distance from s: this produces a line embedding $g_s : V \to \mathbb{R}_+$. For integer $i \ge 0$ define $r_i = \frac{i\cdot\delta}{2\log k}$. Define $\alpha_0 = \text{vol}(V)/k$ and for $i \ge 1$ let $\alpha_i = \alpha_0 + \text{vol}(B(s, r_i))$.

Consider any $1 \leq j \leq 2\log k$. We apply Lemma 8 to the embedding g_s and the interval $[r_{j-1}, r_j]$; note that $\ell(e) < \frac{\delta}{2\log k}$ which implies that we can indeed apply the lemma. Also any edge $e \in \cup_{\theta\in[r_{j-1},r_j]}\delta(S_\theta)$ satisfies the property that $g(u) \in [r_{j-2}, r_{j+1}]$ and $g(v) \in [r_{j-2}, r_{j+1}]$ since $\ell(e) < \frac{\delta}{2\log k}$. Thus

$$\begin{aligned}\int_{r_{j-1}}^{r_j} \nu(\delta(B(s,\theta))d\theta &\leq 2 \sum_{v:g_s(v)\in[r_{j-2},r_{j+1}]} \hat{\rho}_v(\mathbf{d}_v) \\ &\leq 2(\alpha_{j+1} - \alpha_{j-2}). \end{aligned} \tag{2.5}$$

We claim that there is some $1 \leq j < 2\log k$ such that $\alpha_{j+1} \leq 8\alpha_{j-2}$. Suppose not, then $\alpha_{3i} > 8\alpha_{3(i-1)}$ for all $1 \leq i \leq \frac{2\log k}{3}$. This implies that $\alpha_{3i} > 8^i\alpha_0 = 2^{3i}\alpha_0$. Therefore, with $i = \frac{2\log k}{3}$, this implies that $\alpha_{2\log k} > 2^{2\log k}\frac{\mathrm{vol}(V)}{k} > 4\mathrm{vol}(V)$ which is impossible.

Thus there exists a j such that $\alpha_{j+1} \leq 8\alpha_{j-2}$. Consider that j, equation (2.5) implies that

$$\begin{aligned}\int_{r_{j-1}}^{r_j} \nu(\delta(B(s,\theta))d\theta &\leq 2(\alpha_{j+1} - \alpha_{j-2}) \\ &\leq 2(7\alpha_{j-2}).\end{aligned}$$

If we pick r uniformly at random from the interval $[r_{j-1}, r_j]$, where satisfies the above property, the expected cost of $\nu(\delta(B(s,r)))$ is

$$\frac{1}{r_j - r_{j-1}}\int_{r_{j-1}}^{r_j} \nu(\delta(B(s,\theta))d\theta \leq \frac{28\log k}{\delta}\alpha_{j-2},$$

from the preceding inequality and the fact that $r_j - r_{j-1} = \frac{\delta}{2\log k}$. Hence there exists an $r \in [r_{j-1}, r_j]$ such that $\nu(\delta(B(s,r))) \leq \frac{28\log k}{\delta}\alpha_{j-2}$. Since $\alpha_{j-2} - \alpha_0 \leq \mathrm{vol}(B(s,r))$, the lemma follows. $\square$

Now we consider the following algorithm for finding a multicut from a given fractional solution.

- Let $F \leftarrow \{e \mid \ell(e) \geq \frac{1}{4\log k}\}$.
- $G' \leftarrow G[E \setminus F]$.
- Until there exists a pair $s_i t_i$ connected in G', do the following:
  - Let $s_j t_j$ be a pair connected in G'.

- Via Lemma 9 with $\delta = 1/2$ find $r < 1/2$ such that $\nu(\delta_{G'}(B_{G'}(s_j, r))) \leq 2a \log k \cdot (\text{vol}(B_{G'}(s_j, r)) + \text{vol}(V)/k)$.
- $F \leftarrow F \cup \delta_{G'}(B_{G'}(s_j, r))$.
- Remove the vertices $B_{G'}(s_j, r)$ and edges incident to them from G'.

• Output F as the multicut.

Lemma 10. *The set of edges F output by the algorithm is a feasible multicut for the given instance.*

Proof. (Sketch) One can prove this by induction on the number of steps in the while loop. We consider the first step. The diameter of the ball $B_{G'}(s_j, r)$ is $2r < 1$ and hence the end points of any pair cannot both be inside this ball. We remove the edges $\delta(B_{G'}(s_j, r))$ and by the preceding observation there is no need to recurse on this ball. The algorithm recurses on the remaining graph $G' - B_{G'}(s_j, r)$, and by induction separates any pair with both end points in that graph. □

Now we argue about the cost of the set F output by the algorithm. Let $F_0 \leftarrow \{e \mid \ell(e) \geq \frac{1}{4 \log k}\}$ be the initial set of edges added to F and let F_i be the set of edges added in the i'th iteration of the while loop.

Lemma 11. $\nu(F_0) \leq 8 \log k \cdot \sum_v \hat{\rho}_v(\mathbf{d}_v)$.

Proof. For $v \in V$ let $A_v = \{e \in \delta(v) \cap F_0 \mid \ell(e, v) \geq \frac{1}{8 \log k}\}$. We can upper bound $\nu(F_0)$ by $\sum_v \rho_v(A_v)$ since the latter term counts each edge $uv \in F_0$ in at least one of A_u and A_v since $\ell(e, u) + \ell(e, v) = \ell(e) \geq \frac{1}{4 \log k}$. From the definition of the Lovász extension

$$\hat{\rho}_v(\mathbf{d}_v) = \int_0^1 \rho_v(\mathbf{d}_v^\theta) d\theta \geq \int_0^{1/(8 \log k)} \rho_v(\mathbf{d}_v^\theta) d\theta \geq \frac{1}{8 \log k} \rho_v(A_v),$$

where we used non-negativity of ρ_v for the first inequality above and monotonicity for the second. □

Lemma 12. $\sum_{i \geq 1} \nu(F_i) \leq 4a \log k \sum_v \hat{\rho}_v(\mathbf{d}_v)$.

Proof. (Sketch) From the algorithm description, $F_i = \delta(B_{G'}(s_j, r))$ for some terminal s_j and radius $r < 1/2$ where G' is the remaining graph in iteration i. Moreover, $\nu(F_i) \leq 2a \log k \cdot (\text{vol}(B_{G'}(s_j, r)) + \text{vol}(V)/k)$. Since the nodes in

$B_{G'}(s_j, r)$ are removed from the graph, a node u is charged only once inside a ball. Hence

$$\sum_i \nu(F_i) \le \sum_i 2a \log k \cdot \mathrm{vol}(V)/k + 2a \log k \sum_v \hat{\rho}_v(\mathbf{d}_v) \le 4a \log k \sum_v \hat{\rho}_v(\mathbf{d}_v),$$

since there are at most k iterations of the while loop; each iteration separates at least one pair. □

Since ν is subadditive (see Lemma 1)

$$\nu(F) \le \nu(F_0) + \sum_{i \ge 1} \nu(F_i) \le (8 + 4a) \log k \sum_v \hat{\rho}_v(\mathbf{d}_v).$$

This finishes the proof of Theorem 8.

2.5.3 Better approximation for sparsest cut under product demand

In this section, we present a better approximation for the sparsest cut problem under a product demand structure. By product demand, we mean that the demand function has the following structure: the demand between node u and node v is given by $D_{uv} = \pi(u)\pi(v)$, where $\pi : V \to \mathbb{R}^+$ is a function on the nodes. The associated cut problem is interesting because it corresponds to finding sparse separators in graphs which in turn can be used to find balanced separators; these have several applications.

The main result in this setting is a smaller $O(\sqrt{\log k})$ gap between a convex program with quadratic constraints and the sparsest cut problem. The key technical machinery is the theorem on embedding a *negative type* metric into a L_1 space by [14], which can be interpreted as a line-embedding; this fact was used by [45] to obtain an $O(\sqrt{\log k})$-approximation for sparsest cut in node-capacitated graphs. Note that this is not a traditional flow-cut gap result since the SDP-based relaxation used is strictly stronger than the dual of the multicommodity flow relaxation.

The basic idea is to relax the integer program by allowing real values and then adding additional constraints in such a way that a tighter bound can be obtained on the approximation gap, while the program still remains convex and thus efficiently computable. We begin by describing this modified program:

$$
\begin{array}{rcl}
\mathcal{P} & := & \min \sum_v \hat{\rho}_v(\mathbf{d}_v) \\
\ell(e,u) + \ell(e,v) & = & \ell(e) \qquad e = uv \in E \\
\sum_{i,j \in V} \pi_i \pi_j \cdot \ell(ij) & = & 1 \\
\ell(pq) + \ell(qr) & \geq & \ell(pr) \quad \forall p,q,r \in V \\
|\vec{x}_u - \vec{x}_v|^2 & = & \ell(uv) \quad \forall u,v \in V \\
\ell(e), \ell(e,u), \ell(e,v) & \geq & 0 \qquad e = uv \in E \\
\vec{x}_u \in \mathbb{R}^n & & \forall u \in V.
\end{array}
$$

Theorem 9. *A $O(\sqrt{\log k})$-approximate solution to the sparsest cut problem can be found in polynomial time, given an optimal solution to $\mathcal{P}$. Furthermore, $\mathcal{P}$ is a convex program for which an ϵ-additive approximation can be computed in polynomial time.*

Proof. Now we proceed to obtain a solution to the sparsest cut problem given a solution to $\mathcal{P}$. Observe that if integrality conditions are enforced on $\ell(e)$ and $\ell(e,u)$, then the program is the same as the sparsest cut problem and thus the optimal value r of $\mathcal{P}$ is less than or equal to the value of the sparsest cut s.

A metric (V,d) is said to be of *negative type*, if the $(V,\sqrt{d})$ is isomorphic to a subset of the Euclidean space. Any feasible assignment of $\mathcal{P}$ yields a negative-type metric (V,ℓ) with $\ell(u,v) = \ell(e)$, where $e = uv$. We need the following lemma on line-embeddings of negative-type metrics.

Lemma 13. *[14] [15] For every n-point metric space (V,d) and every product weight function $w : V \times V \to \mathbb{R}_+$ with $w(u,v) = \pi(u)\pi(v)$, there is a polynomial-time computable contraction $g : V \to \mathbb{R}$ such that avgd$_w(g) = O(\sqrt{\log n})$. Moreover, if the support of π is k there is a map g such that avgd$_w(g) = O(\log k)$.*

We start with the optimal solution of $\mathcal{P}$, whose optimal value is r, and obtain the line embedding g guaranteed in Lemma 13. Without loss of generality, we can assume that g maps V to the interval $[0,\beta]$ for some $\beta > 0$. We show that there is a θ such that $\delta(S_\theta)$ is a cut of sparsity $rO(\sqrt{\log k})$ by averaging over random cuts $\delta(S_\theta)$ corresponding to θ uniform in $[0,\beta]$. This statement is proved by using Lemma 6 and Lemma 7 in exactly the same way as Theorem 6 and is omitted here for brevity. The cut $\delta(S_\theta)$ thus obtained is an approximately good sparse cut with approximation ratio $O(\sqrt{\log k})$.

Now we show that the program $\mathcal{P}$ is convex: define X as a $n \times n$ matrix with $\vec{x}_u$ as the columns, and $Z := X^T X$ (we have $Z_{ij} = \vec{x}_i \cdot \vec{x}_j$). We can then replace X in the program by Z and set Z to be positive semidefinite: $Z \succeq 0$; this is equivalent since any positive semidefinite Z admits a decomposition of the form $Z = X^T X$. Now, the equation defining $\ell(uv)$ can be replaced as follows:

$$\ell(uv) = (\vec{x}_u - \vec{x}_v)^T \cdot (\vec{x}_u - \vec{x}_v) \tag{2.6}$$

$$\iff \ell(uv) = Z_{uu} + Z_{vv} - Z_{uv} - Z_{vu}. \tag{2.7}$$

Under this new parametrization, the objective function is convex in the variables $\ell(e), \ell(e,u), Z$ and the constraints are also convex. Thus $\mathcal{P}$ is a convex program. This program can be solved efficiently to within an additive ϵ error in time polynomial in $\log(\frac{1}{\epsilon})$ by using the ellipsoidal algorithm. $\square$

2.6 Throughput Flow-Multicut Gaps in Planar and Minor-free Graphs

In this section, we consider the flow-cut gap in undirected *planar* polymatroidal networks[4] and more generally networks (equivalently graphs) that exclude the the complete graph K_h as a minor for some fixed h. For these networks, Gupta et al. [57] conjectured that the concurrent flow-sparsest cut gap is $O(1)$ for the edge-capacitated setting; Rao [129] proved an upper bound of $O(\sqrt{\log n})$ thereby improving upon the gap for general graphs which can be $\Omega(\log n)$ in the worst case. The throughput flow-multicut gap is however known to be $O(1)$ [140] in such graphs. Much less is known for node-capacitated planar graphs; the only result that we are aware of is that of Brinkman, Karagiozova and Lee [20] that shows an $O(\sqrt{\log n})$ gap for series-parallel graphs. Our main result gives an $O(1)$ bound for the throughput flow-multicut gap in minor-free polymatroidal graphs.

Before we proceed to stating our main result formally, we will recall some standard definitions. A graph H is called a *minor* of a graph G if H can be obtained by G by a sequence of edge deletions, vertex deletions and contraction of edges (i.e., collapsing two nodes connected by an edge into a single node). A family of graphs $\mathcal{G}$ is said to be H-minor free, if H is not a minor of G for any

[4] By a planar polymatroidal network we simply mean that the underlying graph G is planar.

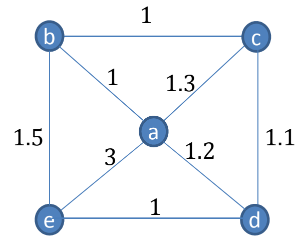


Figure 2.4: Example of a weighted graph

$G \in \mathcal{G}$. Observe that, if $\mathcal{G}$ is H-minor free, $\mathcal{G}$ is also $K_{|H|}$-minor free (where K_h is the complete graph on h nodes).

We will show that for any family of graphs that exclude a fixed minor H, the gap between the maximum throughput flow and minimum multicut is a constant (which depends only on $|H|$). This shows that the gap is a constant for planar graph families using Kuratowski's theorem [84], which states that the set of planar graphs is precisely the graph family that excludes K_5 and $K_{3,3}$ (the complete 3×3 bipartite graph).

We state our main result now, which shows a constant flow-cut gap for any family of graphs that excludes a minor of size.

Theorem 10. *Given a multicommodity problem on a graph $G \in \mathcal{G}$ with polymatroidal constraints, the minimum multicut is within a factor $O(h^2)$ of the maximum concurrent flow if $\mathcal{G}$ is K_h-minor free.*

As an easy corollary we obtain the following result.

Corollary 2. *There is an $O(h^2)$-approximation for finding a minimum* node-weighted *multicut in a graph that excludes K_h as a minor.*

The rest of this section is dedicated to proving Theorem 10. We start with an optimal solution to the relaxation given by edge lengths $\ell(e), e \in E$ and the partition of $\ell(e)$ into variables $\ell(e, u)$ and $\ell(e, v)$, where $e = uv$. We will show that there exists a multicut $F \subseteq E$, which separates each source from its corresponding sink satisfying $\nu(F) = O(h^2) \sum_v \hat{\rho}_v(\mathbf{d}_v)$.

Chopping operation

We will describe a chopping operation, which is used to partition the network. We use the terminology of [89] to describe this process.

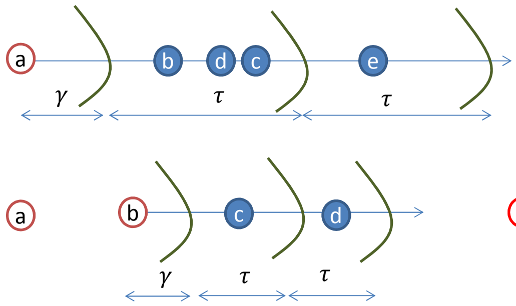


Figure 2.5: Applying two τ-chops iteratively

Given a connected graph H, a special node $v_0 \in V(H)$, positive numbers τ and γ, and a metric ℓ on the nodes, we define a partitioning operation, called τ-chop of H rooted at v_0 with offset γ, as follows. Consider a line embedding of the nodes $V(H)$, induced by the shortest path distance from v_0 using the metric ℓ; i.e., $g : V \to \mathbb{R}_+$ is defined as

$$g(u) = \text{dist}_\ell(u, v_0) \quad \forall u \in V(H). \tag{2.8}$$

Since the graph is connected, $g(u)$ is bounded, and therefore define

$$g_{\max} = \max_{u \in V(H)} g(u). \tag{2.9}$$

The τ-chop partitioning operation divides V into partitions V_i defined as follows:

$$V_i = \{v \in V(H) : \gamma + (i-1)\tau \leq g(v) < \gamma + i\tau\}, \quad i = 1, 2, .., \frac{\lceil g_{\max} \rceil}{\tau}. \tag{2.10}$$

Clearly $V(H) = \biguplus_i V_i$. This partitioning operation disconnects the edges,

$$F := \{e = uv \in E(H) : \exists i \neq j \text{ s.t. } u \in V_i, v \in V_j\}. \tag{2.11}$$

Thus we can think of F as the cut associated with the τ-chop. The cost of the τ-chop is equal to the cut cost $\nu(F)$.

More generally, a τ-chop on a disconnected graph is defined as the result of performing a τ-chop on each of its connected components. When we perform a sequence of τ-chops, the i-th chop performs partitioning individually on each of the partitions created by the $i-1$-th chop.

An example of a graph with distance function is shown in Fig. 2.4. The application of two successive τ-chops to this network is shown in Fig. 2.5. In this figure, the root node for the chop is shown in a transparent red circle, whereas the other nodes are shown as filled blue circles. Observe that in each iteration, for each connected component, we use a distinct line embedding depending upon the root node selected.

We will show that there exists a "good" offset γ, such that the cost of the cut is within a constant factor of the dual cost.

Lemma 14. *Given a graph $G = (V, E)$, a distance metric ℓ satisfying $\ell(e) \leq \tau\ \forall e \in E$, any root node $v_0 \in V$ and a positive number τ, let the offset γ be uniformly random in $[0, \tau]$ and F_γ be the random cut corresponding to the τ-chop rooted at v_0 with offset γ. Then the expected value of the random cut F_γ is*

$$\mathbb{E}\left[\nu(F_\gamma)\right] \leq \frac{2}{\tau} \sum_v \hat{\rho}_v(\mathbf{d}_v). \tag{2.12}$$

Proof. We consider the case when the graph is comprised of a single connected component. The case of a disconnected (partitioned) graph can be dealt with by dealing with each of the connected components (partitions) separately.

We begin by considering the line embedding $g(u)$ induced by the shortest path distance from v_0 using distances ℓ, i.e., $g(u) = \text{dist}_\ell(u, v_0), \quad \forall u \in V(H)$. The length of edge $e = uv$ in the embedding is given by $\ell'(e) := |g(v) - g(u)| \leq \ell(e)$ where we have used the triangle inequality. While the cost $\nu(F_\gamma)$ is in general complicated to evaluate, we can upper-bound $\nu(F_\gamma)$ by using one particular way to assign every edge $e = uv \in F_\gamma$ to either of the nodes u or v, i.e., by charging the edge to the submodular constraint on either u or v. To do this, we critically use the finer grain information contained in the dual variables, $\ell(e, u)$ and $\ell(e, v)$, which add up to give $\ell(e)$ in the relaxation. Let $r = \frac{\ell(e,u)}{\ell(e)}$ and let $\ell'(e, u) = r\ell'(e)$ and $\ell'(e, v) = (1 - r)\ell'(e)$. For $g(u) < g(v)$, we partition the interval $[g(u), g(v)]$ into $[g(u), g(u) + \ell'(e, u))$ and $[g(u) + \ell'(e, u), g(v)]$. If the τ-chop cuts the edge $e = uv$ in the former interval, we assign it to u , else we assign it to v, i.e., charge it to the submodular constraint at v. This is illustrated in Fig. 2.6, where the cut is charged to node v.

In order to state a formal bound on $\nu(F_\gamma)$, we need some definitions. Consider a node u and let $L_u = \{uv \in \delta(u) \mid g(v) < g(u)\}$ be the set of edges uv that go from u to the left of u in the embedding g. Similarly $R_u = \{uv \in \delta(u) \mid g(v) \geq$

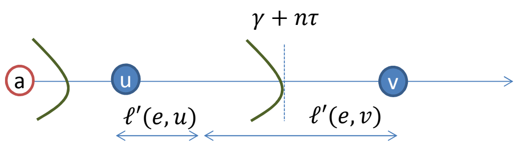


Figure 2.6: Charging an edge

$g(u)\}$. Note that L_u and R_u partition $\delta(u)$. Let $\mathbf{d}'_u$ be the vector of dimension $|\delta(u)|$ consisting of the values $\ell'(e,u)$ for $e \in \delta(u)$. We obtain $\mathbf{d}^L_u$ from $\mathbf{d}'_u$ by setting the values for $e \in R_u$ to 0 and similarly $\mathbf{d}^R_u$ from $\mathbf{d}'_u$ by setting the values for $e \in L_u$ to 0. Since $0 \leq \ell'(e,u) \leq \ell(e,u)$ for each $e \in \delta(u)$ we see that $\mathbf{d}'_u \leq \mathbf{d}_u$ and (component wise) and hence $\mathbf{d}^L_u \leq \mathbf{d}_u$ and $\mathbf{d}^R_u \leq \mathbf{d}_u$. Since ρ_u is monotone, the extension $\hat{\rho}$ is also monotone and we have $\hat{\rho}_u(\mathbf{d}^L_u) \leq \hat{\rho}_u(\mathbf{d}_u)$ and $\hat{\rho}_u(\mathbf{d}^R_u) \leq \hat{\rho}_u(\mathbf{d}_u)$.

We start with

$$\mathbb{E}\left[\nu(F_\gamma)\right] = \frac{1}{\tau}\int_{\gamma=0}^{\tau} \nu(F_\gamma), \tag{2.13}$$

and upper-bound $\nu(F_\gamma)$, for any fixed γ using the assignment formalized below. Define

$$A^L_{\gamma,u} = \left\{ \begin{array}{ll} e = uv \in L_u : \exists n \in \mathbb{N}: & g(u) > \gamma + n\tau \\ & g(u) - \ell'(e,u) < \gamma + n\tau \end{array} \right\}, \tag{2.14}$$

and similarly define

$$A^R_{\gamma,u} = \left\{ \begin{array}{ll} e = uv \in R_u : \exists n \in \mathbb{N}: & g(u) < \gamma + n\tau \\ & g(u) + \ell'(e,u) > \gamma + n\tau \end{array} \right\}. \tag{2.15}$$

Since $\ell(e) \leq \tau$, we note that $A^L_{\gamma,u} \cup A^R_{\gamma,u} = \emptyset$, i.e., only one of the sets is active for a given γ. We can write the upper bound on $\nu(F_\gamma)$ using these sets as

$$\nu(F_\gamma) \leq \sum_{u \in V} \rho_u(A^L_{\gamma,u}) + \rho_u(A^R_{\gamma,u}). \tag{2.16}$$

For a fixed node u, let us consider $\mathbb{E}\left[\rho_u(A^L_{\gamma,u})\right]$. To compute this, let us order the set L_u as $L_u = \{e_1, ..., e_h\}$ such that $\ell'(e_1,u) \geq \ell'(e_2,u) \geq \ldots \geq$

$\ell'(e_h, u) \geq 0$. When we take a random $\gamma \in [0, \tau]$, the probability that edge e_i is cut and assigned to node u, is given by $\frac{\ell'(e,u)}{\tau}$. Furthermore, we observe that whenever edge e_i is cut and assigned to u, all the edges $e_1, ..., e_i$ are also cut and assigned to u. Thus the set of edges $e_1, ..., e_i$ is assigned to node u with probability $\frac{\ell'(e_i,u)-\ell'(e_{i-1},u)}{\tau}$. This gives us the equality

$$\mathbb{E}\left[\rho_u(A^L_{\gamma,u})\right] = \frac{1}{\tau}\sum_{j=1}^{h}(\ell'(e_j, u) - \ell'(e_{j-1}, u))\rho(\{e_1, e_2, \ldots, e_j\}). \tag{2.17}$$

By the definition of Lovász extension, the right-hand side of this equation is equal to $\frac{1}{\tau}\hat{\rho}_u(\mathbf{d}^L_u)$. We can perform a symmetric calculation for the expected value of the second term in (2.16),

$$\mathbb{E}\left[\rho_u(A^L_{\gamma,u})\right] = \frac{1}{\tau}\hat{\rho}_u(\mathbf{d}^L_u). \tag{2.18}$$

Thus (2.16) implies that

$$\begin{aligned} \mathbb{E}\left[\nu(F_\gamma)\right] &\leq \frac{1}{\tau}\sum_{u\in V}\hat{\rho}_u(\mathbf{d}^L_u) + \hat{\rho}_u(\mathbf{d}^R_u) && (2.19)\\ &\leq \frac{2}{\tau}\sum_{u\in V}\hat{\rho}_u(\mathbf{d}_u), && (2.20) \end{aligned}$$

where the second inequality follows due to the fact that $\hat{\rho}(\mathbf{d}^L_u) \leq \hat{\rho}(\mathbf{d}_u)$ and $\hat{\rho}(\mathbf{d}^R_u) \leq \hat{\rho}(\mathbf{d}_u)$. □

We use the following lemma from [76, 43, 89] that shows that if a graph excludes K_h as a minor, then a sequence of $h-1$ τ-chops will yield components with diameter $O(h\tau)$. The lemma below is the formulation in [89].

Lemma 15. *[76, 43, 89] If $G = (V, E)$ with distances $\ell(e), e \in E$ excludes K_h as a minor, then for any $\tau \geq 1$, any sequence of $h-1$ iterated τ-chops on V results in a partition $V = S_1 \cup S_2 \cup \cdots \cup S_m$ such that* $\mathrm{diam}(S_i) \leq O(h\tau)$*, where* diam *refers to the diameter in G using the shortest path distance dist$_\ell$.*

2.6.1 Algorithm for finding a multicut

- Compute the optimal solution to the relaxation. This can be done efficiently using the ellipsoidal algorithm, since the separation oracle for the dual is a simple shortest path computation.

- Initialize $F \leftarrow F_0 := \{e \mid \ell(e) \geq \tau\}$, i.e., remove all edges greater than length τ.
- Set $G' \leftarrow G[E \setminus F]$ with distance function $\ell(e), e \in E(G')$.
- Perform $(h-1)$ τ-chops on G' as follows. For the i-th chop, choose an arbitrary node in each connected component as the corresponding root node and use uniformly independently chosen offsets $\gamma \in [0, \tau]$. Let F_i be the cut associated with the i-th τ-chop. For each $i = 1, 2, ..., h-1$, update

$$F \leftarrow F \cup F_i. \tag{2.21}$$

- Output F as the multicut.

Since the graph avoids K_h as a minor, by Lemma 15, the diameter of every component will be smaller than $O(h\tau)$. By setting $\tau = \frac{\triangle}{Ch}$, with C large enough, the diameter of every component will be smaller than $\triangle$. We set $\triangle = \frac{1}{2}$, which implies that for any i, s_i and t_i are never in the same component due to the fact that $\text{dist}_\ell(s_i, t_i) \geq 1$ and the triangle inequality which implies that $\text{dist}_\ell(s_i, v_0) + \text{dist}_\ell(t_i, v_0) \geq 1$.

Theorem 11. *The algorithm outputs a multicut F such that*

$$\mathbb{E}\left[\nu(F)\right] \leq O(h^2) \sum_v \hat{\rho}_v(\mathbf{d}_v). \tag{2.22}$$

Proof. We will compute the cost of the multicut F as follows:

$$\mathbb{E}\left[\nu(F)\right] \leq \nu(F_0) + \sum_{i=1}^{h-1} \mathbb{E}\left[\nu(F_i)\right], \tag{2.23}$$

since the cost function $\nu(.)$ is subadditive (this follows from the fact that ρ_v is a polymatroid, and hence is subadditive).

We first compute the cost $\nu(F_0)$ as follows. Since for each edge $e = uv \in E$, $\ell(e) \geq \tau$, either $\ell(e, u) \geq \frac{\tau}{2}$ or $\ell(e, v) \geq \frac{\tau}{2}$ as $\ell(e) = \ell(e, u) + \ell(e, v)$. Define for $v \in V$, $A_v = \{e \in \delta(v) \cap F_0 \mid \ell(e, v) \geq \frac{\tau}{2}\}$. We can upper-bound $\nu(F_0)$ by $\sum_v \rho_v(A_v)$ since the latter term counts each edge $uv \in F_0$ in at least one of A_u or

A_v. From the definition of the Lovász extension,

$$\hat{\rho}_v(\mathbf{d}_v) = \int_0^1 \rho_v(\mathbf{d}_v^\theta)d\theta \geq \int_0^{\tau/2} \rho_v(\mathbf{d}_v^\theta)d\theta \geq \frac{\tau}{2}\rho_v(A_v),$$

where we used non-negativity of ρ_v for the first inequality above. The second inequality follows from the fact that $A_v \subseteq \mathbf{d}_v^\theta$, whenever $\theta \leq \frac{\tau}{2}$ and the monotonicity of ρ_v. Thus, we get

$$\nu(F_0) \leq \sum_v \rho_v(A_v) \leq \frac{2}{\tau}\sum_v \hat{\rho}_v(\mathbf{d}_v). \tag{2.24}$$

By Lemma 14, we get that, for the i-th τ-chop, the expected cost is

$$\mathbb{E}\left[\nu(F_i)\right] \leq \frac{1}{\tau}\sum_v \hat{\rho}_v(\mathbf{d}_v). \tag{2.25}$$

Substituting this into (2.23), we get

$$\begin{aligned} \mathbb{E}\left[\nu(F\right] \quad &\leq \quad \frac{h+1}{\tau}\sum_v \hat{\rho}_v(\mathbf{d}_v) = \frac{Ch(h+1)}{\triangle}\hat{\rho}_v(\mathbf{d}_v) && (2.26) \\ &= \quad O(h^2)\sum_v \hat{\rho}_v(\mathbf{d}_v), && (2.27) \end{aligned}$$

using the choice $\triangle = \frac{1}{2}$, which concludes the proof of the theorem. $\square$

Clearly Theorem 11 implies Theorem 10 and we are done.

Proof of Corollary 2: A multicut in a node-weighted graph G can therefore be modeled by a multicut in a polymatroidal network G' obtained from G as follows. For each v with weight $w(v)$ we define the function ρ_v as : $\rho_v(S) = w(v)$ for each $S \subseteq \delta(v)$, $S \neq \emptyset$. Note that the multicut in the polymatroidal network G' is defined as a set of edges F but its cost $\nu(F)$ takes into account the minimum weight set of nodes whose removal ensures that all edges of F are removed. For instance if an edge $uv \in F$ is assigned to u in the evaluation of $\nu(F)$ then the node u will be part of the multicut in the original graph G.

CHAPTER 3

MULTIPLE UNICAST IN WIRELESS NETWORKS

"The wireless telegraph is not difficult to understand. The ordinary telegraph is like a very long cat. You pull the tail in New York, and it meows in Los Angeles. The wireless is the same, only without the cat." – attributed to Albert Einstein

3.1 Background

Having studied multiple-unicast in the polymatroidal network model, we now turn our attention to the main focus of this thesis: studying multiple unicast traffic in wireless networks. As stated in the introduction, there are two distinct objectives:

- To obtain the approximate capacity of multiple unicast in wireless networks, and
- To establish a layered communication architecture that can guide engineering design.

We will accomplish these two objectives simultaneously by constructing layered communication strategies that are near optimal for multiple unicast in general wireless networks.

3.1.1 Organization

The rest of this chapter is organized as follows:

- In Sec. 3.1.2, we describe prior work and its relation to the work in this chapter.
- An overview of the layering approach is provided in Sec. 3.2. After defining the "locality" over which local physical layer schemes must be implemented, a list of desirable properties of local solutions are provided.

- In Sec. 3.3, polymatroidal networks, which form the backbone of the layering architecture, are reviewed.
- In Sec. 3.4, “good” local physical layer solutions are described for various channel models. For some models, it is shown that existing schemes satisfy the desirable properties, whereas for other models, where existing schemes are insufficient, new ones are constructed.
- In Sec. 3.5, the local schemes are fitted into a global network context. Capacity theorems are proved for the various channel settings by connecting the wireless network problem formally to the polymatroidal network problem.

3.1.2 Prior work

Fundamental understanding of layering architectures has recently received plenty of attention from the networking community [27, 134], and scenarios have been identified under which a joint optimization of the transport and network layers naturally decompose into separate optimization problems, thus yielding a justification for the layered architecture. While there have been attempts to include certain aspects of the wireless medium into this framework [113], the understanding is far from complete. In this thesis, we take a fundamental, information theoretic perspective on if and when the physical, medium access and network layers can be separately designed.

Capacity results for wireless networks: Substantial progress has been made in the recent past in understanding the key aspects of the wireless medium (broadcast and superposition) from an information-theoretic view point. In particular, the capacity of MIMO broadcast channel has been resolved [153], approximate capacity of the 2-user interference channel has been established [40], the approximate capacity of 2-user X-channels characterized [100] [62], and the degrees of freedom of K-user interference channels with diversity characterized (beginning with the seminal work of [21], several papers have strengthened this result [112, 39, 107]). While these results establish information theoretic understanding of several important (“physical layer”) channels, there is no conceptual guideline to fit the solutions for reliable communication for the channel in the context of a bigger network it could be a part of. In a different direction, there has been significant progress in understanding network-level capacity issues in the context of simple traffic models, starting from the breakthrough work [17], where the approximate capacity of single unicast is characterized, and later generalized to several scenarios: the approximate capacity of unicast in discrete memoryless networks is characterized in [91], a separation result between analog and digital components in relay networks is established in [9] and the approximate capacity of broadcasting in Gaussian networks is established in [69]. More recently, it has been demonstrated that simple amplify-and-forward schemes can achieve the approximate capacity of unicasting in Gaussian networks under certain conditions [101].

While unicast traffic in general Gaussian networks and multiple unicast traffic in single-hop Gaussian networks are reasonably well understood, the capacity of multiple unicast traffic in Gaussian wireless networks remains an open problem in multi-terminal information theory. In recent times, several research groups have made progress on this problem [137, 105, 64, 145, 11, 58] , but the general problem still remains unsolved. Specific directions, with promise of success, involve simplifying the problem by considering specific traffic patterns such as 2-unicast [105, 137, 83, 151, 152, 66, 139, 53, 147]; another approach is to consider more specific network topologies, like for example, K sources communicating to K sinks via L fully-connected layers of K relays each [64, 117]. While these existing works attempt to compute the degrees-of-freedom (or approximate capacity) exactly for specific instances of the problem, we adopt a different viewpoint and focus our attention on obtaining general results for arbitrary networks (at the expense of obtaining potentially weaker approximation in specific instances).

In the context of multiple-unicast in large wireless networks, there has been significant progress in understanding *scaling laws* for geographical wireless networks, beginning with the seminal work in [55] and culminating in the hierarchical relaying scheme in [121] and a combination of the two [115, 116, 120] (with several critical works in between [156, 96, 1, 157, 46]). Despite its significant advantages, the performance guarantees are only in the context of certain specific wireless network models and, more importantly, the communication scheme is not a representation of a simple layered architecture for communication.

Information-theoretic layering architectures: Separation theorems form a basic tool in information theory: in his celebrated paper [135], Shannon showed that source coding (compression) and channel coding (communication) can be separated without loss of optimality. Following this, several separation (and non-separation) theorems have been proved in the multi-terminal context (see, for example, [32, 141]). The result most relevant to the current discussion is the separation between network coding and channel coding proved in the pioneering work in [80]. There it is shown that, for a wireline network composed of *independent noisy channels*, a separation architecture composed of a physical layer that performs independent coding for each channel and a network layer which transports bits across the induced noiseless network, is optimal. This is a very interesting structural result that holds under arbitrary traffic models and is proved without the necessity to compute the capacity of the network. Thus the question of studying the capacity of wireline networks can be reduced to the question of studying the capacity of capacitated *graphs*.

For k-unicast in wireline graphs, a very interesting dichotomy is known in the theoretical computer science literature: for undirected graphs, the classical work of Leighton and Rao [90] shows that routing achieves the min-cut to within a $\mathcal{O}(\log k)$ factor and, furthermore, there is a standing conjecture [92, 59, 78, 86] which claims that routing is in fact optimal; in contrast, for general directed graphs, it has been shown recently [28] that unless P=NP, there is no polynomial time algorithm that can approximate the value of min-cut to within a k^ϵ factor for some $\epsilon > 0$. Since max-flow can be computed in polynomial time, this result implies that there are networks for which flow cut gap is greater than k^ϵ. Furthermore, it has been proved recently [22] that computing the network coding region for directed graphs is equivalent to determining the entropy vector region, which is believed to be a very hard problem. Thus the string of positive results in the context of undirected (or bidirected) graphs and negative results in the context of directed graphs serve as an indicator that it may be easier to understand bidirected wireless networks.

A study of layering in directed wireless networks is initiated in [81]. The key idea is that of channel emulation, where a given channel is upper-bounded by a wireline network with joint capacity constraints in such a way that the wireline network can emulate all possible behavior of the channel. This is a very strong condition, which ensures that the channel can be upper-bounded by the wireline network *irrespective of the traffic pattern*. This program has been already accomplished for networks of 2-user MAC and broadcast channels, but appears to be very hard for general networks. In this thesis, we instead work with only multiple-unicast traffic but go on to study layering in general bidirected networks.

Polymatroidal networks: Polymatroidal networks and approximation theorems for these networks form a critical component in this thesis. For detailed background and prior work on polymatroidal networks, we refer the reader to Chapter 2. Here we survey some applications of polymatroidal networks in the information theory literature. The polymatroidal structure of the multiple access channel capacity region was observed and exploited by Tse and Hanly [142]. Directed polymatroidal networks were utilized in the work of Vasudevan and Korada [146], where a separation architecture for a network of deterministic broadcast and MAC channels converts the network into a polymatroidal network and existing results for broadcasting in polymatroidal networks [44] are used to obtain capacity bounds. Recently, there have been several applications of network flows, and their generalizations such as flows in linking systems [132], to unicast information flow in wireless networks [17, 8, 160, 52, 124, 75]. It is worth noting that linking systems generalize polymatroidal networks.

3.2 Layered Architecture

Engineering approaches to reliable network communication involve "layering," a separation of the roles of physical (dealing with channel uncertainty), medium access (dealing with sharing the wireless medium) and networking (dealing with end-to-end the resulting "wireline" network communication). On the other hand, fundamental architectures are suggested by information theoretic study of large wireless networks (a major research direction in the past decade, with performance measured in a coarse scaling context). For instance, multihop routing [55] is a layered architecture, while hierarchical MIMO [121] (nearly scaling-law optimal in a geographically uniform context) is not. The information theoretic understanding of layering architectures has recently started receiving attention (see [80, 81]). Our approach is along similar lines as the approach in [146], where a layered architecture for a network of deterministic broadcast and MAC channels is used to obtain capacity bounds.

In understanding the systematic design of layered architectures, it helps to look at the global wireless network as a collection of "local" wireless networks. The focus of this section is to introduce this viewpoint; we propose that the notion of locality comes from both geographic (spatial) and temporal contexts. We see that certain combinatorial properties of the (physical layer) solutions to the local networks are *desirable*; these will help prove fundamental guarantees on the performance of the layered architecture in the global context.

3.2.1 Locality

A wireless network is a collection of *local* channels, if there are no interactions between the channels. Formally, a wireless network is defined as follows: Consider a graph $G = (V, E)$. For each $v \in V$, x_v, y_v denote the transmit and received symbol respectively. Let $C^+(v), C^-(v)$ denote the set of all channels in which v can transmit and receive from respectively, i.e., x_v can be written as $x_v = \{x_v^c\}_{c \in C^+(v)}$, and $y_v = \{y_v^c\}_{c \in C^-(v)}$. We will consider a wireless network with independent noise, where

$$\mathbb{P}(y_1, \ldots, y_n | x_1, \ldots, x_n) = \Pi_v \mathbb{P}(y_v | \{x_u\}_{u:(u,v) \in E}), \tag{3.1}$$

and this description explicitly captures the relationship between the graph and the joint probability transition function.

Consider a set of channels $c \in \mathcal{C}$. A wireless network is said to be composed of channels $c \in \mathcal{C}$ if the probabilistic description of the network is of the form

$$\begin{aligned} \mathbb{P}(y_1, \ldots, y_n | x_1, \ldots, x_n) &= \Pi_c \mathbb{P}(\{y_v^c\}_{v \in V^-(c)} | \{x_u^c\}_{u \in V^+(c)}) & (3.2) \\ &= \Pi_c \Pi_{v \in V^-(c)} \mathbb{P}(y_v^c | \{x_u^c\}_{u \in V^+(c):(u,v) \in E}), & (3.3) \end{aligned}$$

where $V^-(c) = \{v : c \in C^-(v)\}$ and $V^+(c) = \{v : c \in C^+(v)\}$. Each channel c is referred to as a component channel of the network.

A canonical scenario occurs when the wireless network is simply a collection of statistically independent noisy channels. Here each channel between a transmitter and a receiver is local. A more interesting example occurs in the case of a *frequency planned* wireless network, where each component of the wireless network operates in a specified frequency range. Here, the overall channel model can be decomposed as the product of channel models in each frequency range; the scale of locality corresponds to the scale of frequency reuse.

In general, such a geographic decomposition (via frequency planning) may not happen. Nevertheless, we can view the decomposition as occurring in *time* (indeed, this has been a popular method for analyzing general wireline / wireless networks [4, 17]). When we decompose across time, the local channel corresponds to the global one, as viewed over a specific single block of time. In this context, the layering architecture restricts the sophistication of physical layer (and medium access layer) strategies to be restricted to operate on a single layer in time, and at the end of each epoch, the information is *decoded* and re-encoded (using the networking layer) for the next local channel. The layering architecture thus enforces decoding of all information at each "hop" (in time); schemes, such as quantize-and-forward [17], which forward analog information do not fit the layered architecture model.

3.2.2 Desirable properties of local solutions

A natural desirable property of any (physical layer) solution to a local channel is it be as optimal (from an information theoretic view point) as possible. In particular, we will be interested in how close the solution is to fundamental upper bounds given by the cutset bounds and certain natural combinatorial properties of the solution. For a network described by a probability transition matrix, the cut-set bound can be written as follows. Given a cut $\Omega \subseteq V$, let D_Ω be the set of demands separated by the cut, i.e., $D_\Omega := \{k : s_k \in \Omega, t_k \in \Omega^c\}$. The cut-set bound bounds the sum of rates of sources in D_Ω and can be written as[1]

$$\sum_{k \in D_\Omega} R_k \leq \sup_{p_{x_1,\dots,x_n}} \mathrm{I}(X_\Omega; Y_{\Omega^c} | X_{\Omega^c}). \tag{3.4}$$

Our focus on cut-set bounds as opposed to specialized outer bounds for specific wireless channels (such as the broadcast and interference channels) is due to the following reasons.

- *Generality*: The cut-set bound [38] is an information theoretic outer bound on the achievable rate region and it can be written down for a general wireless network.
- *Decomposition*: The chain rule of mutual information allows the cut-set bound of a network to decompose into the cut-set bounds on local channels; thus solutions that come close to the cut-set bound at a local level have a potential to be layered and be still close to the cut-set bound at a global

[1]While there is a stronger way of writing this bound, this weaker form of the bound will suffice for the purposes here.

level. Formally, if we have a cut $\Omega \subseteq V$, the value of the cut is given by

$$\begin{aligned}
\mathrm{I}(X_\Omega; Y_{\Omega^c}|X_{\Omega^c}) &= H(Y_{\Omega^c}|X_{\Omega^c}) - H(Y_{\Omega^c}|X_V) && (3.5)\\
&= H(Y_{\Omega^c}|X_{\Omega^c}) - \sum_c H(Y^c_{\Omega^c}|X^c_V) && (3.6)\\
&\leq \sum_c H(Y^c_{\Omega^c}|X_{\Omega^c}) - H(Y_{\Omega^c}|X_V) && (3.7)\\
&\leq \sum_c \mathrm{H}(Y^c_{\Omega^c}|X^c_{\Omega^c}) - \mathrm{H}(Y_{\Omega^c}|X_V) && (3.8)\\
&= \sum_c \mathrm{H}(Y^c_{\Omega^c}|X^c_{\Omega^c}) - \mathrm{H}(Y_{\Omega^c}|X^c_V) && (3.9)\\
&= \sum_c \mathrm{I}(X^c_\Omega; Y^c_{\Omega^c}|X^c_{\Omega^c}), && (3.10)
\end{aligned}$$

and thus the cut-set decomposes into sum of the cut-sets evaluated for each channel.

- *Structure*: Cut-set bounds have been well studied in the theoretical computer science literature and their combinatorial structure has been well understood. In fact, algorithms for approximately computing the cut-set bounds form an integral part of the theory of approximation algorithms.

- *Invariance under feedback*: The cut-set bound (evaluated under general joint distributions) is essentially obtained by upper bounding the rate of communication of the separated sources by the rate of a point-to-point channels and is, therefore, invariant to feedback.

Finally, *reciprocity* of the local channels (rate region reciprocity with the roles of transmitter and receiver reversed) will be paid attention to. The combinatorial structure imposed by the bidirected nature of each local channel will yield to efficient algorithms that are close to cuts.

3.2.3 Layering methodology

Layering architectures stitch together the local solutions into a global solution:

1. The solution to a local channel allows for reliable digital communication at a local level.

2. Replacing each local channel by a set of (wireline) links leads to a network comprised of noiseless channels, with the rates on the various edges being *coupled* by the rate regions of the local solutions. Reciprocal local solutions ensure that the network obtained is *bidirected* (i.e., any edge between node a and node b has a corresponding edge between node b and a, with the two edges being involved in the same types of capacity region constraints).

3. Over the resulting wireline network, we might have to employ network coding to re-encode the information between local channels. We utilize the combinatorial properties of the coupled rate constraints to study this new class of wireline networks. In particular, if the combinatorial structure governing the rate constraints is a specific form of a polytope known as a *polymatroid*, we obtain polymatroidal networks. Therefore, we study polymatroidal networks (which have local polymatroidal constraints on rate region) and prove that routing can achieve the cut-set bound to within a $\mathcal{O}(\log k)$ factor for the k-unicast problem, and also prove some better approximations for more specific communication problems.

4. Since the cut-set bound on a network of channels decomposes into a *sum* of cut-set bounds on the local channels, we can readily compare the performance of the layering architecture to a fundamental upper bound on the global network performance.

Whenever local solutions are close to the cut-set bounds for the corresponding local channels, we can establish the fundamental near-optimality of the layering architecture. We have accomplished this program for several canonical local wireless channels including broadcast erasure channels, Gaussian uplink and downlink channels, and interference channels with diversity (example: fast fading).

3.3 Polymatroidal Networks

In order to keep this chapter self-consistent, we provide a quick overview of polymatroidal networks. For further details, we refer the reader to Chapter 2.

3.3.1 Polymatroids

A set function $f : 2^N \to \mathbb{R}$ over a finite ground set N is submodular iff $f(A) + f(B) \geq f(A \cap B) + f(A \cup B)$ for all $A, B \subseteq N$; equivalently $f(A \cup \{i\}) - f(A) \geq f(B \cup \{i\}) - f(B)$ for all $A \subseteq B$ and $i \notin A$. It is monotone if $f(A) \leq f(B)$ for all $A \subset B$. A polymatroid refers to the following set in $\mathbb{R}^N$:

$$\mathcal{P} = \left\{ (x_1, ..., x_N) : \sum_{i \in S} x_i \leq f(S) \quad \forall S \subseteq [N] \right\}, \tag{3.11}$$

where $f(S)$ is a monotone submodular function with $f(\emptyset) = 0$. Thus a polymatroid is fully specified by specifying a monotone submodular function with $f(\emptyset) = 0$ (we will call such a function itself as a polymatroid). An example of a polymatroid is the following: given a set of N vectors $v_1, ..., v_N$ in $\mathbb{R}^M$, the function $f(S)$ defined as the rank of the matrix composed of $\{v_i\}_{i \in S}$ defines a polymatroid (we refer the reader to [118] for an introduction on polymatroids) .

3.3.2 Definition of polymatroidal networks

A commonly studied wireline scenario is one where each edge is labeled by a capacity: this is the largest amount of information flowing on that edge. Here we are interested in a more general model which is able to handle the additional constraints when edges meet at a node, similar in spirit to the broadcast and superposition constraints in wireless.

Consider a node v in a directed graph G and let $\delta_G^-(v)$ be the set of edges in to v and $\delta_G^+(v)$ be the set of edges out of v. In the standard model each edge (u, v) has a non-negative capacity $c(u, v)$ that is independent of other edges. In the (directed) polymatroidal network for each node v there are two associated submodular functions: ρ_v^- and ρ_v^+ which impose joint capacity constraints on the edges in $\delta_G^-(v)$ and $\delta_G^+(v)$ respectively. That is, for any set of edges $S \subseteq \delta_G^-(v)$, the total capacity available on the edges in S is constrained to be at most $\rho_v^-(S)$, similarly for $\delta_G^+(v)$. Note that an edge (u, v) is influenced by ρ_u^+ and ρ_v^-. For the definition of flow and cut in polymatroidal networks, we refer the reader to Sec. 2.2.

We define a bidirected polymatroidal network as a directed polymatroidal network with the following properties.

- Every edge $e = (i, j)$ has a corresponding reverse edge $\tau(e) := (j, i)$.
- For any vertex v, the polymatroidal constraint $\rho_v^-(\cdot)$ on the incoming edges $\text{In}(v)$ is the same as the polymatroidal constraint $\rho_v^+(\cdot)$ on the outgoing edges $\text{Out}(v)$. More concretely,

$$\rho_v^-(E_v) = \rho_v^+(\tau(E_v)) \quad \forall E_v \subseteq \text{In}(v). \tag{3.12}$$

3.3.3 Main result

The following theorem is proved in Sec. 2.5, which generalizes the results of [90, 97] to the case of polymatroidal capacity networks:

Theorem 12. *For a bidirected polymatroidal network with k source-destination pairs, the ratio between the max-flow rate region and the min-cut rate region is $\mathcal{O}(\log k)$. The max-flow and an approximate min-cut can be calculated in polynomial time. Furthermore, this factor is tight in general, i.e., there are families of polymatroidal networks such that the flow-cut gap is $\Omega(\log k)$.*

3.3.4 Special traffic scenarios

While in general, the factor of $\mathcal{O}(\log k)$ for flow-cut gaps is tight for multiple-unicast in bidirected polymatroidal networks, there may be special communication scenarios for multiple unicast when the factor can be improved. We present some instances here, where the flow cut gap is much better even for the more general case of *directed* polymatroidal networks.

Broadcast traffic: Broadcast traffic is a special type of multiple unicast traffic where all the messages originate at a single source. Consider a directed polymatroidal network with a single source s having independent messages to K destination nodes $t_1, ..., t_K$.

Lemma 16. *[44] For a directed polymatroidal network with broadcast-traffic pattern, the rate region of the max flow equals the rate region of the min-cut.*

Sum rate in directed X networks: Consider J sources S_1,...,S_J and K destinations T_1,...,T_K, where each source has an independent message for each destination. The rate tuple is a JK length vector R_{jk} between each j and k. This communication problem is referred to commonly as the X-network problem.

Lemma 17. *For a directed polymatroidal network with X-traffic pattern, the* sum-rate *of max flow equals the sum-rate bound given by min-cut.*

Proof. Construct a super source S which talks to the J sources with infinite capacity links and a super sink T which is connected from each of the K sinks via infinite capacity links. The max-flow min-cut theorem for unicast between S and T in directed polymatroidal networks [88] implies the desired result. □

Sum rate for group communication in directed networks: Consider a directed polymatroidal network with a specially marked group of nodes $S \subseteq V$. Each node s in S has an independent message for every other node in S. Thus it is a multiple unicast problem with $|S|(|S|-1)$ messages. We refer to this traffic pattern as the group-communication traffic pattern. Suppose we are interested only in maximizing the sum-rate.

Lemma 18. *For a directed polymatroidal network with a group-communication traffic pattern, the sum-rate of max-flow is greater than* half *the sum-rate bound given by min-cut.*

Proof. The proof of this theorem is non-trivial and requires a reduction from the directed polymatroidal network to the directed edge capacitated network using a combinatorial uncrossing argument. For a directed edge-capacitated network, this theorem is proved by Naor and Zosin [109]. For a detailed proof of this statement for polymatroidal networks, we refer the reader to Remark 1 in Chapter 2. □

3.4 Local Physical Layer Schemes

In this section, good local physical layer schemes for several channel models will be discussed. For each of these channels, the goals will be to identify a physical layer scheme, quantify its rate region, understand its closeness to the cut-set bound and to examine its combinatorial structure. We will also analyze if the rate region remains (approximately) the same when the sources and destinations are exchanged and channels are reversed. In later sections, these properties will allow us to stitch together local physical layer schemes to get global schemes. The results in this section for various channel models are summarized in Table 3.1.

We will use the notation $\mathcal{R}^{\text{ch}}_{\text{ach}}$ and $\mathcal{R}^{\text{ch}}_{\text{cut}}$ to denote the achievable and the cut-set *rate regions* respectively, where the superscript 'ch' denotes the channel of interest.

Table 3.1: Canonical wireless channels, as viewed via three lenses.

Characteristic / Channel	Closeness-to-Cut	Combinatorial Structure	Reciprocity
Linear Deterministic MAC / BC	Exact	Polymatroidal	Exact
Gaussian MAC / BC	Approximate	Polymatroidal	Approximate
Erasure Broadcast	Far	Polytope	Far
Erasure Broadcast (Feedback)	Approximate	Polymatroidal	Approximate
Fading MAC / BC with delayed CSI	Approximate	Polymatroidal	Approximate
Fading linear deterministic	Approximate	Polymatroidal	Approximate
Fading X-Channel	Approximate	Polymatroidal	Approximate
Fixed X-Channel	Approximate in DOF	Polymatroidal DOF region	Approximate in DOF

3.4.1 Linear deterministic broadcast and multiple access channels

Consider a broadcast channel with d receivers of the form

$$y_i = H_i x \ , \quad \forall i = 1, 2, ..., d, \tag{3.13}$$

where y_i is the received vector at receiver i and x is the transmitted vector. The source intends to communicate independent messages to each of its destinations. For a subset K of $\{1, 2, ..., d\}$, let H_K denote the matrix with $H_i, i \in K$ stacked up alongside one another. The capacity region of this broadcast channel [103] is given by

$$\mathcal{C} = \{(R_1, ..., R_d) : \sum_{i \in K} R_i \leq \text{Rank}(H_K) \quad \forall K \subseteq [d]\}. \tag{3.14}$$

This capacity region is also equal to the cut-set bound, which is a polymatroid (see [118]).

Let us consider a "reciprocal" multiple access channel in which there are d transmitters and one receiver,

$$y = \sum_i H_i^T x_i, \tag{3.15}$$

and all the transmitters have an independent message to transmit to the single destination. The capacity region of a general MAC channel is known (see, for example, Chapter 14 in [31]) and for the linear deterministic channel, is given again by the rate region in (3.14). We observe that the capacity of the broadcast channel and the reciprocal MAC channel are the same and are equal to their cut-set bound.

Thus for a linear deterministic broadcast and MAC channel, the rate region is exactly polymatroidal, equal to the cut, and is reciprocal.

3.4.2 Gaussian broadcast and multiple access channels

Let us first consider a multiple access channel, defined by

$$y = \sum_i h_i x_i + z, \tag{3.16}$$

where the transmitted vector x_i is constrained by a power constraint P at each of the d nodes, y is the received vector and z denotes the noise, which is of unit power. Let the rate region achievable on this multiple access channel be denoted by $\mathcal{R}_{\text{ach}}^{\text{MAC}}(P)$. This region is known to be polymatroidal [142]. This rate region equals the cut-set bound evaluated under product distributions [31] , i.e.,

$$\mathcal{R}_{\text{ach}}^{\text{MAC}}(P) = \mathcal{R}_{\text{cut,product}}^{\text{MAC}}(P). \tag{3.17}$$

Let the cut-set bound evaluated under general distributions be given by $\mathcal{R}_{\text{cut,general}}^{\text{MAC}}(P)$. We can easily verify the relation

$$\mathcal{R}_{\text{ach}}^{\text{MAC}}(P) \supseteq \mathcal{R}_{\text{cut,general}}^{\text{MAC}}\left(\frac{P}{d}\right). \tag{3.18}$$

Next, let us consider a "reciprocal" or "dual" broadcast channel, given by

$$y_i = H_i x + z_i \ , \quad \forall i = 1, 2, ..., d, \tag{3.19}$$

where the transmitted vector x is constrained by a power constraint dP, y_i is the received vector and z_i denotes the noise at each receiver, which is of unit power. Let us call the rate region of the broadcast channel as $\mathcal{R}^{\text{BC}}(P)$. This rate region has been fully characterized, but is not equal to the cut-set bound and is not polymatroidal (see Chapter 6 in [143] for a discussion).

The rate region of this broadcast channel with sum power constraint Pd contains the rate region of the multiple access channel with *individual constraint* P at each node [149, 148], i.e.,

$$\mathcal{R}_{\text{ach}}^{\text{BC}}(P) \supseteq \mathcal{R}_{\text{ach}}^{\text{MAC}}(P). \tag{3.20}$$

For the purpose of symmetry, we can choose to operate the broadcast channel at the rate region specified by $\mathcal{R}^{\text{MAC}}(P)$, i.e., let us set

$$\mathcal{R}_{\text{ach}}^{\text{BC}}(P) = \mathcal{R}_{\text{ach}}^{\text{MAC}}(P). \tag{3.21}$$

This will also ensure that the rate region of the achievable scheme $\mathcal{R}_{\text{ach}}^{\text{BC}}(P)$ is polymatroidal. Thus, in our achievable strategy, the rate region of a multiple access and that of the dual broadcast channel are equal and given by a polymatroidal region.

Let the cutset bound for the broadcast channel be specified as $\mathcal{R}^{\text{BC}}_{\text{cut,general}}(P)$. Since there is only one input variable, the cut-set bound under product distribution and general distribution are the same.

Lemma 19. *The achievable region of the MAC channel compares with the cutset bound under general distributions as follows:*

$$\mathcal{R}^{MAC}_{ach}(P) \supseteq \mathcal{R}^{MAC}_{cut,general}\left(\frac{P}{d}\right). \tag{3.22}$$

For the broadcast channel, we have the relation

$$\mathcal{R}^{BC}_{ach}(P) \supseteq \mathcal{R}^{BC}_{cut,general}\left(\frac{P}{d}\right). \tag{3.23}$$

Proof. The proof is deferred to Appendix B.1. □

Thus for a Gaussian broadcast and MAC channel, the rate region is approximately polymatroidal, close to the cut, and is approximately reciprocal.

3.4.3 Broadcast erasure channels

Consider a network comprised of broadcast erasure channels. For a broadcast erasure channel with d receivers, the channel model can be written as

$$y_i = e_i x, \quad \forall i = 1, 2, ..., d, \tag{3.24}$$

where e_i is a binary random variable which when 0 represents that at receiver i, the packet got erased. If e_i are all independent, then the broadcast erasure channel is said to be an independent erasure broadcast channel. Such a channel is specified by erasure probabilities $\epsilon_i, i = 1, 2, ..., d$, where

$$\epsilon_i = \text{Pr}\{e_i = 0\}. \tag{3.25}$$

For the purpose of simplicity in this thesis we will consider only broadcast erasure channels that are independent and symmetric, which implies that there is only one parameter ϵ and $\epsilon_i = \epsilon \quad \forall i$.

Lemma 20. *For an erasure broadcast channel without feedback, the cut-set bound can be as large as L times the achievable rate region.*

Proof. See Appendix B.2. □

This result implies that for broadcast erasure channels (without feedback), there are no good local schemes that can achieve close to the cut.

Broadcast erasure channels with feedback: Since there are no good local schemes for the broadcast channel, we suggest that the scale of the locality be enlarged to include the presence of feedback links in the physical model to get better local schemes.

The capacity of the erasure broadcast channel with ACK feedback (the receiver acknowledges whether it received the packet) was studied by [50] for the two user channel, and later extended to the more general case independently by [150] and [49]. The schemes are based on network coding and interference alignment and demonstrate that the following rate region is achievable.

Lemma 21. *[150, 49] The following rate region is achievable for the erasure broadcast channel with ACK feedback:*

$$R_{\mathit{ach,fb}} = \left\{ (R_1, ..., R_D) | \sum_{i=1,2,..,d} \frac{R_{\pi(i)}}{1-\epsilon^i} \leq 1 \quad \forall \pi \right\}. \tag{3.26}$$

Here π is a permutation of the set $(1, \ldots, d)$. Note that this region is not polymatroidal. However, it is close to the min-cut rate region (which is itself polymatroidal) as seen below:

Lemma 22.

$$\mathcal{R}_{\mathit{ach,fb}} \supseteq \frac{\mathcal{R}_{\mathit{cut}}}{\mathcal{O}(\log d)}. \tag{3.27}$$

Proof. See Appendix B.3. □

The reciprocal nature of wireless channels from which the broadcast erasure channel is constructed naturally suggests a way of providing feedback links of *commensurate* strength. Formally: a channel is said to have *commensurate feedback* if there are feedback links from the various receiving nodes to the transmitters with the same rate region as the cut-set bound for the forward channel. In Appendix B.4, we look at one possible way of obtaining feedback links of commensurate strength as the forward links.

Thus for a broadcast erasure channel with feedback, the rate region is approximately polymatroidal, close to the cut, and is approximately reciprocal.

3.4.4 MIMO broadcast and MAC channel with delayed feedback

We will consider MIMO broadcast and MAC channels, similar to Sec. 3.4.2, but with the difference that the channel states are i.i.d. over antennas and time. If the CSI is instantaneously available at the transmitter and the receiver, then the methods and results of Sec. 3.4.2 continue to hold. However, the problem becomes interesting when global channel state information is no longer available. In some settings, the channel may change so fast that by the time the channel state feedback reaches the transmitter, the channel state has changed significantly. This feature is essentially present in the erasure channel if we think of erasure as being a channel state, which is unknown to the transmitter originally. In the erasure channel case, the ACK feedback delivers this channel state (i.e., whether the channel is in erasure or not) to the transmitter with a delay. Recent work by Ali and Tse [7] showed the surprising result that, even in Gaussian networks, delayed CSI can be very beneficial as compared to the absence of CSIT [159]. This will form the basis of our investigation of these channels.

We assume that each broadcast and MAC channel gets feedback from its receivers about the channel state; however, the channel changes before the feedback arrives, precluding the use of feedback to predict the future state of the channel. We will resort to a degree-of-freedom characterization for this problem. Let d_i denote the achievable degrees of freedom for the i-th message, i.e.,

$$d_i := \lim_{\mathtt{SNR}\to\infty} \frac{R_i(\mathtt{SNR})}{\log \mathtt{SNR}}, \tag{3.28}$$

where R_i is the achievable rate for user i. The achievable DOF region is denoted by $\mathcal{D}_{\text{ach}}$, and the region given by the cutset bound is denoted by $\mathcal{D}_{\text{cut}}$.

Such a multiple access case is easy to deal with because even without channel state information, the cut-set bound can be achieved,

$$\mathcal{D}_{\text{ach}}^{\text{MAC}} = \mathcal{D}_{\text{cut}}^{\text{MAC}}. \tag{3.29}$$

However for the corresponding broadcast channel, in the absence of CSIT, we can only achieve rates that are far from the cut-set. In [7], it is shown that the presence of channel state feedback, even when delayed, can significantly alter the situation, as was the case with broadcast erasure channels:

Lemma 23. *[7] For a fading MISO broadcast channel with a source with L trans-*

mit antennas and L single transmit antenna receivers, the following DOF region can be achieved with the help of delayed CSIT:

$$\mathcal{D}_{ach}^{BC} = \left\{ \sum_{i=1}^{L} D_i \leq l \frac{K}{\sum_{i=1}^{L} \frac{1}{i}} \right\}. \tag{3.30}$$

Our first result is to obtain an approximation for the rate region of MIMO broadcast channels with delayed feedback, formally stated in the following lemma.

Lemma 24. *For a fading MIMO broadcast channel with l transmit antennas and d users with user i having m_i antennas, and delayed CSI feedback, the following DOF region can be achieved:*

$$\mathcal{D}_{ach}^{BC} \supseteq \frac{\mathcal{D}_{cut}^{BC}}{\mathcal{O}(\log p)}, \tag{3.31}$$

where $p := \min(l, \sum_i m_i)$ is the minimum of the number of transmit and receive antennas in the system.

Thus for a fading MIMO broadcast channel, the rate region is approximately polymatroidal, close to the cut, and is approximately reciprocal.

3.4.5 Fading X-channels

Consider an L-user X-channel where there are L sources $s_1, ..., s_L$ and L destinations $t_1, ..., t_L$ with *each* source having message to send to each destination. In this channel, there are L^2 messages in total. The connectivity graph between the nodes is a bi-partite graph with E being the set of edges. We will abbreviate edge (s_i, t_j) as (i, j) since the meaning is clear from the context.

Note that an interference channel is a special case of this channel. Capacity achieving schemes even for the L-user interference channel are not known in the general setting. In [21], the authors show that each user in an L-user interference channel can achieve half their point-to-point degrees-of-freedom (DOF) if the channel is fast-fading, using a mechanism called interference alignment, which was initially proposed for the 2-user X-channel in [100]. This result has since been generalized in several directions; most notably, in [112], it is shown that using an "ergodic interference alignment" scheme, each user can get half her rate at *all SNR*, and in [107], it is shown that the DOF result can be proved even under *fixed channel coefficients* using a scheme termed "real interference alignment." These results have been unified into a single framework in [155]. It has also been shown [114] that ergodic interference alignment can be used to achieve linear capacity scaling in dense interference networks.

Cuts in an X-channel:

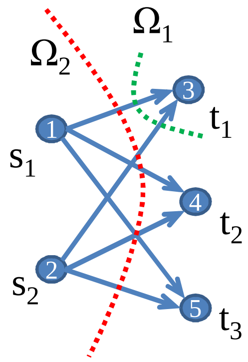


Figure 3.1: Cuts in X-channel

Consider an example X-channel with two sources and three destinations, shown in Fig. 3.1. Two cuts are marked in the figure. The light (green) cut Ω_1 separates only destination 1 from the two sources, thus providing a bound $R_{11}+R_{21} \leq C_{\Omega_1}$, whereas dark (red) cut Ω_2 separates all sources from all destinations and therefore provides a bound on $R_{11} + R_{12} + R_{13} + R_{21} + R_{22} + R_{23} \leq C_{\Omega_2}$. Of these two bounds, the first corresponds to that of a polymatroidal constraint whereas the second does not correspond to a polymatroidal constraint (since in a polymatroidal network, only edges that meet at a node have a joint constraint).

For a general X-channel, these two types of cuts will be present, and we can classify them as

1. Cuts that separate a single node from the rest of the nodes (referred to as cuts of the polymatroidal form), and
2. Cuts that separate multiple sources from multiple destinations.

We would like to show not only an achievable scheme that achieves the cut-set bound approximately, but also that the rate region of the achievable scheme satisfies a polymatroidal constraint. Therefore, for an X-channel, we will have to show that only cuts of the polymatroidal form (separating one node from the rest) play a dominant role. This is a key challenge that we address in this section.

Channel model: The channel model can therefore be written as

$$y_i(t) = \sum_{j \in \text{In}(i)} h_{ij}(t)x_j(t) + z_i(t) \quad \forall t = 1, 2, ..., T, \tag{3.32}$$

where $x_i(t), y_i(t), z_i(t)$ are the transmitted vector, received vector and noise vector at time t, In(i) represents the set of neighbors of node i who have an incoming edge to i and fading coefficient $h_{ij}(t)$ is associated with edge $(i, j) \in E$ at time t. The noise vector is assumed to have unit variance and is independent at each node. There is a power constraint of P per node.

We will make the following assumptions about the fading distribution:

- Fading coefficients are assumed to be i.i.d. over edges and over time.
- The fading coefficient will be assumed to be symmetric, i.e., if h_{ij} is a discrete random variable,

  $$\Pr\{h_{ij} = a\} = \Pr\{h_{ij} = -a\}, \quad \forall a; \tag{3.33}$$

  otherwise, if the random variable is absolutely continuous, the pdf $p(.)$ must satisfy

  $$p(h_{ij} = a) = p(h_{ij} = -a) \quad \forall a. \tag{3.34}$$

- The fading distribution is assumed to satisfy:

  $$a := e^{-\mathbb{E}(\log |h|^2)} < \infty. \tag{3.35}$$

One example of a fading distribution that satisfies these assumptions is when $h_{ij}(t)$ is i.i.d. across nodes and time with a complex Gaussian distribution, for which $a = 1.723$ [119].

We will use the shorthand $C(P)$ to denote the ergodic capacity of a fading channel with power constraint P, and the fading coefficient h of unit variance,

$$C(P) := \mathbb{E}_h \left[\frac{1}{2} \log(1 + |h|^2 P) \right]. \tag{3.36}$$

Scheme for the K-user interference channel: First, we consider the case of k-user interference channel, where there are messages only from s_i to t_i for each $i = 1, 2, ..., k$. In the ergodic set-up, the following result is known:

Lemma 25. *[112] For a k-user ergodic interference channel, where the direct links are non-zero, the following rate tuple is achievable:*

$$R_i = \frac{1}{2}C(2P), \quad \forall i = 1, 2, ..., k. \tag{3.37}$$

Scheme for the X-channel: We generalize this physical layer scheme to the X-channel with L sources and M destinations, demonstrating not only that the cut-set bound is approximately achievable, but also that only cut of the polymatroidal form are relevant.

Theorem 13. *For an L-source, M-sink ergodic X-channel, the following rate region is achievable:*

$$\mathcal{R}_X^{ach} = \left\{ (R_{ij})| \begin{array}{ll} \sum_{j:(i,j)\in E} R_{ij} \leq \frac{1}{2}C(2P) & \forall i \in S \\ \sum_{i:(i,j)\in E} R_{ij} \leq \frac{1}{2}C(2P) & \forall j \in T \end{array} \right\}. \tag{3.38}$$

and furthermore, if d is the maximum degree of any node,

$$\mathcal{R}_{ach}^X(P) \supseteq \frac{\mathcal{R}_{cut}^X(\frac{2P}{ad})}{2}, \tag{3.39}$$

where $a := e^{-\mathbb{E}(\log |h|^2)}$.

Proof. Let us write R_{ij} for the rate of communication between s_i and t_j. We use the following achievable strategy:

- Let us construct a bipartite graph between the source vertices and sink vertices, and edges given by E.
- A matching in a bipartite graph is a choice of edges such that each node is present in at most one edge. In our case, a matching can be thought of as representing a choice of at most one destination for each source. Choose a matching M on the bipartite graph, and let π be the corresponding permutation. The characteristic vector of a bipartite matching is given by the vector $(x_{ij}) : x_{ij} = 1$, if $(i, j) \in M$, otherwise $x_{ij} = 0$.

- Consider the interference channel from $s_1, ..., s_L$ to $d_{\pi(1)}, ..., d_{\pi(L)}$. For the $s_i, t_{\pi(i))}$ pairs that are connected, we can achieve a rate of $\frac{1}{2}C(2P)$ using the strategy of Lemma 25.

- This implies that a rate given by $\frac{1}{2}C(2P)$ times the characteristic vector of the bipartite matching is achievable.

- Now, we can achieve any convex combination of the rates given by matchings on the graph. This is given by the following polytope, called the matching polytope:

$$\mathcal{M} = \text{conv}\{(x_{ij})_M | M \text{ a matching }\}. \tag{3.40}$$

- By a theorem in bi-partite graph matchings [131], this matching polytope can be alternately described as:

$$\mathcal{P} = \left\{ (x_{ij}) | x_{ij} \geq 0 \quad \forall i, j, \begin{array}{cc} \sum_{j:(i,j)\in E} x_{ij} \leq 1 & \forall i \\ \sum_{i:(i,j)\in E} x_{ij} \leq 1 & \forall j \end{array} \right\}. \tag{3.41}$$

- Therefore the achievable rate region is given by (3.38).

- The cut-set bound $\mathcal{R}_{\text{X}}^{\text{cut}}$ implies the following, which are only a subset of the cuts (the cuts which separate one node from all the others) :

$$\left\{ (R_{ij}) | \begin{array}{cc} \sum_{j:(i,j)\in E} R_{ij} \leq \mathbb{E} \log(1 + \sum_{j:(i,j)\in E} h_{ij}^2 P) & \forall i \in S \\ \sum_{i:(i,j)\in E} R_{ij} \leq \mathbb{E} \log(1 + \sum_{i:(i,j)\in E} h_{ij}^2 P) & \forall j \in T \end{array} \right\}.$$

- Now, due to the concavity of the logarithm,

$$\begin{aligned} \mathbb{E} \log(1 + \sum_{j:(i,j)\in E} h_{ij}^2) &\leq \log(1 + \sum_{j:(i,j)\in E} \mathbb{E} h_{ij}^2 P) && (3.42) \\ &\leq \log(1 + dP) && (3.43) \\ &\leq \mathbb{E} \log(1 + adP|h|^2) = C(adP) && (3.44) \end{aligned}$$

where the last step follows because of the convexity of the function $f(x) = \log(1 + ce^x)$, i.e., applying Jensen inequality for the aforementioned convex function, we get

$$\begin{aligned}\mathbb{E}(\log(1+c|h|^2)) &= \mathbb{E}\{\log(1+ce^{\log|h|^2})\} && (3.45)\\ &\geq \log(1+ce^{\mathbb{E}(\log|h|^2)}) && (3.46)\\ &= \log(1+ca^{-1}), && (3.47)\end{aligned}$$

where $a := e^{-\mathbb{E}(\log|h|^2)})$.

Thus the cut-set bound implies the following inequalities:

$$\left\{(R_{ij})| \begin{array}{ll}\sum_{j:(i,j)\in E} R_{ij} \leq C(adP) & \forall i \in S\\ \sum_{i:(i,j)\in E} R_{ij} \leq C(adP) & \forall j \in T\end{array}\right\}.$$

- Therefore we get the result that

$$\mathcal{R}^{\text{X}}_{\text{ach}}(P) \supseteq \frac{\mathcal{R}^{\text{X}}_{\text{cut}}(\frac{2P}{ad})}{2}. \tag{3.48}$$

□

Thus for a fading X channel, the rate region is exactly polymatroidal, approximately close to the cut, and exactly reciprocal (since the description of the rate region remains the same even the channel is reversed).

3.4.6 Fixed X-channels

Consider an L-user X-channel with fixed channel coefficients drawn from a continuous distribution. We will obtain a degrees-of-freedom characterization of this X-channel (which holds almost surely). The channel model can therefore be written as

$$y_i(t) = \sum_{j\in\text{In}(i)} h_{ij}x_j(t) + z_i(t) \quad \forall t = 1, 2, ..., T, \tag{3.49}$$

where $x_i(t), y_i(t), z_i(t)$ are the transmitted vector, received vector and noise vector at time t, In(i) represents the set of neighbors of node i who have an incoming edge to i and channel coefficient h_{ij} associated with edge $(i, j) \in E$ is drawn from a continuous distribution which has a probability density function, i.e., the probability measure is absolutely continuous with respect to the Borel measure.

The noise vector is assumed to have unit variance and is independent at each node. There is a power constraint of P per node.

First, we consider the case of an L-user interference channel, where there are messages only from s_i to t_i for each i. The following result characterizes the degrees of freedom of the L-user interference channel:

Lemma 26. *[107] For a k-user interference channel with channel coefficients drawn from a continuous distribution, if the direct links are non-zero, the following DOF tuple is achievable almost surely:*

$$D_i = \frac{1}{2}, \quad \forall i = 1, 2, ..., k. \tag{3.50}$$

We can generalize this interference channel scheme to the X-channel with L sources and M destinations, using the same method as in Theorem 13.

Theorem 14. *For an L-source, M-sink fixed X-channel, the following DOF region is achievable* almost surely*:*

$$\mathcal{D}_X^{ach} = \left\{ (D_{ij}) | \begin{array}{ll} \sum_{j:(i,j)\in E} D_{ij} \le \frac{1}{2} & \forall i \in S \\ \sum_{i:(i,j)\in E} D_{ij} \le \frac{1}{2} & \forall j \in T \end{array} \right\}. \tag{3.51}$$

and furthermore,

$$\mathcal{D}_{ach}^X \supseteq \frac{\mathcal{D}_{cut}^X}{2} \quad a.s. \tag{3.52}$$

Proof. The proof proceeds in a manner quite similar to that of Theorem 13; the only difference is that we use Lemma 26 instead of Lemma 25. Also, since we are dealing with DOF, which is an SNR-scaling characterization, the (constant) power scaling factor is not relevant. □

Thus for a fixed X-channel, the achievable DOF region is exactly polymatroidal, approximately close to the cut, and exactly reciprocal (since the description of the achievable DOF region remains the same even the channel is reversed).

3.4.7 Fading linear deterministic channels

In Sec. 3.4.1, we considered linear deterministic networks which had only broadcast and MAC components. In this section we will consider a general linear deterministic network with fading. The communication network is represented by a directed graph $G = (V, E)$. If $(i, j) \in E \iff (j, i) \in E$, we call the network bidirected. The edges (i, j) that are present have fading matrix $H_{ij}(t)$ on them. Each fading coefficient in each matrix is distributed i.i.d. fading over edges and time. The fading distribution for each non-zero coefficient is assumed to be uniform over the finite field (except zero) $\mathbb{F}_q$ 0. The proof can be extended to the case where the fading takes value 0 but there will be penalty factor of $\frac{q-1}{q}$.

For a linear deterministic interference network with fading, there is a scheme based on ergodic interference alignment that achieves half the point-to-point rate for each user:

Lemma 27. *[112] For an k-user fading linear deterministic interference channel, with direct links being non-zero, the following rate type is achievable:*

$$R_i = \frac{1}{2}\log_2(q), \quad \forall i = 1, 2, ..., k. \tag{3.53}$$

We can use this scheme to create a scheme for the X-channel, the rate region of this scheme is quantified in the following theorem.

Theorem 15. *For a L-source, M-destination deterministic ergodic X-channel, the following rate region is achievable:*

$$\mathcal{R}^X_{ach} \supseteq \frac{\mathcal{R}^X_{cut}}{2}. \tag{3.54}$$

Furthermore, only cuts that separate one node from the rest are sufficient.

Proof. The proof for this case is very similar to the proof of Theorem 13, except that the term $\frac{1}{2}C(2P)$ is now replaced by $\frac{1}{2}\log_2(q)$. Also, there are no power scaling losses in the case of a linear deterministic network. $\square$

Thus for a linear deterministic X channel with fading, good local schemes exist; furthermore the schemes have a rate region which can be described by polymatroidal constraints.

3.5 Approximate Capacity Results for Wireless Networks

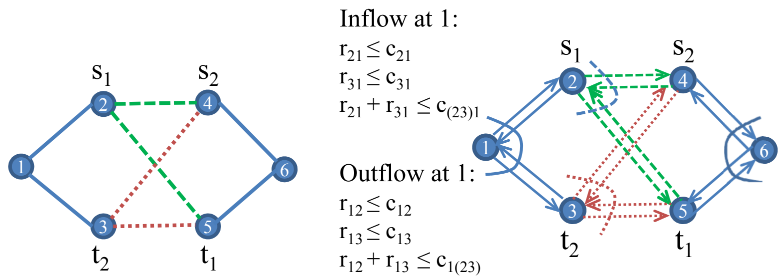
In this section, we will present approximate capacity results for wireless networks with *multiple unicast traffic* for several channel models. Our achievable schemes will use a layering architecture and we will use the cut-set bound as the outer bound. For each channel model, we will have the description of the wireless network as a graph $G = (V, E)$. There are k designated source nodes $s_1, ..., s_k$, which have independent messages and k corresponding destination nodes $t_1, ..., t_k$: t_i wants to decode the message of s_i with vanishingly small probability of error, for every $i = 1, \ldots k$. Let $\mathcal{R}^{\text{ch}}_{\text{ach}}$ denote the rate region comprising the rate tuples achievable and let $\mathcal{R}^{\text{ch}}_{\text{cut}}$ denote the rate region corresponding to the cut-set bound for the channel model ch. If a power constraint P is present, we will describe these regions as a function of P, i.e., $\mathcal{R}^{\text{ch}}_{\text{ach}}(P)$ and $\mathcal{R}^{\text{ch}}_{\text{cut}}(P)$.

3.5.1 Gaussian networks with MAC and broadcast components

We consider Gaussian networks in which there are only broadcast and MAC components, i.e., there is no interference channel component. This is equivalent to the assertion that each edge is involved in either a superposition constraint or in a broadcast constraint but not in both. In practice, such a network can be realized by using a partial frequency reuse scheme, where the total bandwidth is divided into different chunks, which are assigned to users in such a way that interference component is avoided.

Gaussian networks with MAC and broadcast components alone have been previously considered in [81], where it has been shown that for a network composed of 2-user MAC and broadcast channels, a separation architecture involving local coding is approximately optimal. Since this general structural result holds for all possible traffic models, it may be necessary in general to use network coding [4] [79] at the network layer. In contrast, in this thesis, we will assume multiple unicast traffic and utilize the *reciprocity* of the wireless network to show that the cut-set bound can be approximately achieved using a separation scheme along with *routing*.

Network Model: The communication network is represented by an undirected graph $G = (V, E)$, and an edge coloring $\psi : E \to C$, where C is the set of colors. Each node v has a set of colors $C(v) \subseteq C$ on which it transmits and receives. Each color can be thought of as an orthogonal resource (for example, a frequency band), and therefore the broadcast and superposition constraints for



(a) Example of frequency-planned network

(b) Equivalent noiseless network

Figure 3.2: A network composed of broadcast and MAC components and its polymatroidal equivalent

the wireless channel apply only *within* a given color. We will assume the nodes are equipped with full duplex radios on each of these resources. For simplicity of notation, we will assume that there is a one-to-one correspondence between colors and channels, i.e., each channel operates on a distinct color; so c stands for a unique color and a unique channel.

The channel model can therefore be written as

$$y_i^c = \sum_{j \in \mathcal{N}_c(i)} h_{ij}^c x_j^c + z_i^c \quad \forall c \in C(i), \tag{3.55}$$

where x_i^c, y_i^c, z_i^c are the transmitted vector, received vector and noise vector on color c, $h_{ij}^c = h_{ji}^c$ is the channel coefficient between node i and node j on color c and $\mathcal{N}_c(i)$ represents the set of neighbors of node i who are operating on color c and $d_c(i) = |\mathcal{N}_c(i)|$ be the degree of node i in color c. Let $d = \max_{c,v} d_c(v)$ be the maximum degree of any node in a given color; therefore, d is the maximum number of users on any component broadcast or multiple access channel. Each node has a power constraint P *per edge*. Therefore node v has power constraint $Pd_c(v)$ for transmitting on color c. By the very definition, this network has a reciprocal MAC channel for every broadcast channel and vice versa. Let $V(c) = \{v : c \in C(v)\}$ be the set of nodes that use the color c. An example of a wireless network along with its equivalent noiseless network is shown in Fig. 3.2.

Theorem 16. *For the k-unicast problem in a Gaussian network composed of broadcast and multiple access channels, a simple separation strategy can achieve* *a rate*

$$\mathcal{R}^{g}_{ach}(P) \supseteq \frac{\mathcal{R}^{g}_{cut}(\frac{P}{d})}{\mathcal{O}(\log k)}. \tag{3.56}$$

This means that the min-cut, scaled down in power by a factor d and in rate by a factor $\mathcal{O}(\log k)$, can be achieved. For the unicast scenario ($k = 1$), we can show using a similar proof that

$$\mathcal{R}^{\mathrm{g}}_{\mathrm{ach}}(P) \supseteq \mathcal{R}^{\mathrm{g}}_{\mathrm{cut}}\left(\frac{P}{d}\right). \tag{3.57}$$

This result is similar to that obtained by [17], except that here it is obtained for the special case of networks composed of broadcast and multiple access channels. The scheme in [17] requires a global physical layer scheme (the "quantize and map" strategy), while for the special case of networks here we show that a simple separation strategy suffices.

Coding scheme: Proof of Theorem 16

The coding scheme is a separation-based (layered) strategy: each component broadcast or multiple access channel is coded for independently creating bit-pipes on which information is routed globally. The achievable scheme used for the MAC and broadcast channel are discussed in detail in Sec 3.4.2. The achievable rate regions for the MAC and broadcast channels described there are polymatroidal, and therefore each multiple access or broadcast channel with d users can be replaced by a set of d bit-pipes whose rates are jointly constrained by the corresponding polymatroidal constraints. Thus we get a polymatroidal network by using this layered strategy; this polymatroidal network is described as follows: for each node v in the original graph, there are several vertices v_c, one for each color $c \in C(v)$. There is an edge between u_c and v_c if $h^c_{uv} \neq 0$, the polymatroidal constraints are given by

$$\rho^-_{v_c}(F_v) = \log(1 + \sum_{u:(u,v)\in F_v} |h^c_{uv}|^2 P) \qquad \forall F_v \subseteq \delta^-(v_c) \tag{3.58}$$

$$\rho^+_{v_c}(F_v) = \log(1 + \sum_{u:(u,v)\in F_v} |h^c_{vu}|^2 P) \qquad \forall F_v \subseteq \delta^+(v_c), \tag{3.59}$$

and the polymatroidal network is bidirected due to the fact that $h^c_{uv} = h^c_{vu}$ and the reciprocity in the rate regions of the MAC and BC channel. Further, there are edges between v_c and $v_{c'}$ of infinite capacity, since these correspond to the same node v in the original graph.

Let us denote by $\mathcal{R}^{\text{poly}}_{\text{ach}}(P)$ and $\mathcal{R}^{\text{poly}}_{\text{cut}}(P)$ respectively the rate region for the flow-based achievable scheme and the cut-set bound region on this induced polymatroidal network. Then we have from Theorem 12 that

$$\mathcal{R}^{\text{poly}}_{\text{flow}}(P) \supseteq \frac{\mathcal{R}^{\text{poly}}_{\text{cut}}(P)}{\mathcal{O}\log k}. \tag{3.60}$$

As an example, the bidirected polymatroidal network induced for the example of Fig. 3.2a is shown in Fig. 3.2b. The submodular constraints are explicitly written down only for node 1, but similar constraints apply at nodes 2, 3 and 6. In this figure, c_{21}, c_{31} and $c_{(23)1}$ represent constraints on the rate of communication from node 2 to 1, the rate of communication from node 3 to 1 and the sum rate from nodes 2 and 3 to 1 respectively.

It is now sufficient to compare the cuts on the polymatroidal and the Gaussian network.

Lemma 28.

$$\mathcal{R}^{g}_{cut}(P) \subseteq \mathcal{R}^{poly}_{cut}(dP). \tag{3.61}$$

Proof. Given a cut F_Ω in the polymatroidal network, we will show that there is a corresponding cut in the Gaussian network, whose value is within a power scaling factor d of the polymatroidal cut. The value of the cut in the polymatroidal network is $\nu(F_\Omega) = \sum_c \nu(F^c_\Omega)$, i.e., the polymatroidal cut breaks up into the sum of the cuts of various colors. The value of cut in a given color $\nu(F^c_\Omega)$ corresponds to a certain polymatroidal constraint in a given MAC or broadcast channel. We need to show that there is a similar cut in the Gaussian network whose value is within a power scaling factor.

As shown in Lemma 19, we have that for each of these channels, the Gaussian cut and the polymatroid representing the achievable scheme are within a power scaling factor of d, and therefore

$$\mathcal{R}^{\text{g}}_{\text{cut}}(P) \subseteq \mathcal{R}^{\text{poly}}_{\text{cut}}(dP), \tag{3.62}$$

since both polymatroidal and Gaussian cuts decompose into sum of individual cuts. $\square$

The achievable rate using the separation strategy is given using (3.60) as

$$\mathcal{R}^{\text{poly}}_{\text{ach}}(P) \supseteq \frac{\mathcal{R}^{\text{poly}}_{\text{cut}}(P)}{\mathcal{O}(\log k)} \supseteq \frac{\mathcal{R}^{\text{g}}_{\text{cut}}(\frac{P}{d})}{\mathcal{O}(\log k)}. \tag{3.63}$$

This completes the proof of Theorem 16.

Special traffic scenarios: We now present results for *directed* networks with MAC and broadcast components under the special traffic patterns presented in Sec. 3.3.4.

Theorem 17. *For a directed Gaussian network composed of broadcast and multiple access channels, a simple separation strategy can achieve a rate*

$$\mathcal{R}_{ach}(P) \supseteq \mathcal{R}_{cut}\left(\frac{P}{d}\right) \quad \textit{for BC Traffic}, \tag{3.64}$$

$$R^{sum}_{ach}(P) \supseteq R^{sum}_{cut}\left(\frac{P}{d}\right) \quad \textit{for X Traffic}, \tag{3.65}$$

$$R^{sum}_{ach}(P) \supseteq \frac{R^{sum}_{cut}\left(\frac{P}{d}\right)}{2} \quad \textit{for group-communication traffic}. \tag{3.66}$$

Proof. The proofs are based on the polymatroidal results Lemma 16, Lemma 17 and Lemma 18. The proof is very similar to the proof of Theorem 16 and is therefore omitted. □

3.5.2 Broadcast erasure networks with commensurate feedback

Broadcast erasure networks, in which there are broadcast but no superposition constraints, serve as high level models for communication in wireless networks. Unicast in broadcast erasure networks is well understood, for which it has been shown [33] that min-cut is achievable using a global linear network coding scheme (in [94], it is shown that knowledge of erasure locations is not necessary at the destination) . It has also been shown that a separation scheme in which each broadcast erasure channel is coded for locally to create noiseless links does not perform very well. This is due to the fact that for each broadcast erasure channel the capacity region is far away from the min-cut region. However, as shown in Sec. 3.4.3, by utilizing ACK feedback, the capacity region is enlarged to become closer to the min-cut region. Therefore we consider a network of broadcast erasure channel, where each channel has a feedback mechanism.

Consider a network which is composed of broadcast erasure channels, with an appropriate mechanism for feedback built into the network. In particular, we look for feedback that is commensurate; formally: a channel is said to have *commensurate feedback* if there are feedback links from the various receiving nodes to the transmitters with the same rate region as the cut-set bound for the forward channel. The reciprocal nature of wireless channels from which the broadcast erasure channel is constructed naturally suggests a way of providing feedback links of commensurate strength. In Appendix B.4, we look at one possible way of obtaining feedback links of commensurate strength as the forward links.

From now on, we will assume that each erasure broadcast channel has commensurate feedback, without cognizance to how this particular rate region was obtained. For simplicity, we will assume that all broadcast erasure channels are symmetric and independent, i.e., each broadcast erasure channel has erasure independent probability ϵ.

Theorem 18. *For the k unicast problem in a network of erasure broadcast channels with commensurate feedback, a simple separation strategy can achieve a rate*

$$\mathcal{R}_{ach}^{erasure} \supseteq \frac{\mathcal{R}_{cut}^{erasure}}{\mathcal{O}(\log k)\mathcal{O}(\log d_{\max})}, \tag{3.67}$$

where $d_{\max}$ is the maximum number of users in any broadcast erasure channel.

Proof. Consider the following separation strategy: even though the feedback links have a rate region given by $R_{\text{cut}}^{\text{BC}}$, we will restrict them to use the rate region $R_{\text{ach,fb}}^{\text{BC}}$ in order to preserve symmetry. The feedback links are used for two distinct purposes:

- To provide Ack / Nack feedback. This feedback has an overhead of 1-bit per packet which we treat as negligible. This assumption makes sense especially when packet lengths are large.

- To route flows on the reverse direction. Since the Ack feedback overhead is assumed to be small, this will essentially occupy the whole capacity. We establish a bidirected network by using the feedback links for routing.

Since we have the Ack / Nack feedback for each erasure broadcast channel, we can use the scheme of Lemma 21 to obtain the rate region $R_{\text{ach,fb}}^{\text{BEC}}$ with feedback. This induces a bidirected polymatroidal network specified in the following manner, in which we can use flows to achieve a rate region $R_{\text{ach}}^{\text{poly}}$. By Theorem 12, we have that

$$R_{\text{ach}}^{\text{poly}} \supseteq \frac{\mathcal{R}_{\text{cut}}^{\text{poly}}}{\mathcal{O}(\log k)}. \tag{3.68}$$

By Lemma 22, we have $\mathcal{R}_{\text{ach,fb}}^{\text{BEC}} \supseteq \frac{\mathcal{R}_{\text{cut}}^{\text{BEC}}}{\mathcal{O}(\log d_{\max})}$. Further, since cuts in the polymatroidal and the original network decompose into cuts for each channel, any cut-set in the polymatroidal network induced by the achievable scheme has a counterpart cut-set in the erasure network within a factor of $\mathcal{O}(\log d_{\max})$:

$$\mathcal{R}_{\text{cut}}^{\text{poly}} = \frac{\mathcal{R}_{\text{cut}}^{\text{erasure}}}{\mathcal{O}(\log d_{\max})}. \tag{3.69}$$

Now (3.68) and (3.69) together imply

$$\mathcal{R}_{\text{ach}}^{\text{poly}} \supseteq \frac{\mathcal{R}_{\text{cut}}^{\text{erasure}}}{\mathcal{O}(\log k)\mathcal{O}(\log d_{\max})}, \tag{3.70}$$

which proves the desired result. □

Special traffic scenarios: We now present results for *directed* networks with broadcast erasure channels under the special traffic patterns presented in Sec. 3.3.4. Since the networks are directed, reciprocity is not needed; however, we will continue to assume that the broadcast erasure channel has ACK feedback.

Theorem 19. *For a directed network composed of broadcast erasure channels with ACK feedback, a simple separation strategy can achieve a rate*

$$\mathcal{R}_{ach} \supseteq \frac{\mathcal{R}_{cut}}{\log(d_{\max}+1)} \quad \textit{for BC Traffic}, \tag{3.71}$$

$$R_{ach}^{sum} \supseteq \frac{R_{cut}^{sum}}{\log(d_{\max}+1)} \quad \textit{for X Traffic}, \tag{3.72}$$

$$R_{ach}^{sum} \supseteq \frac{R_{cut}^{sum}}{2\log(d_{\max}+1)} \quad \textit{for group-communication traffic}, \tag{3.73}$$

where $d_{\max}$ is the maximum degree of the broadcast channel.

Proof. The proofs are similar to the proof of Theorem 18 and are therefore omitted. □

3.5.3 Gaussian fast fading network

We will now consider a general Gaussian network where broadcast and superposition can simultaneously occur, i.e., the network can contain interference components. Such a network is clearly more general than the K-user interference channel, and even this channel has not been well understood in the most general case (the tutorial [61] provides an excellent summary of the current understanding on this channel). However, in the presence of fast fading, the problem gets symmetrized considerably [112], and there is a reasonable understanding of this problem. Therefore we will resort to the fast fading model in this section. We will further assume that the fading distribution satisfies the assumptions in Sec. 3.4.5.

While most of the existing literature is on single-hop interference channels, multi-hop interference networks have been the focus of more recent work. In particular, it has been shown in [63] that the degrees of freedom of such fully-connected layered networks can be achieved using a non-separation scheme called opportunistic interference alignment. It has also been shown [53] that a separation architecture does not even achieve the DOF for simple networks, for example, the network with 2 sources, 2 relays and 2 destinations. Our results offer a contrast: if we look to achieve the capacity within an approximation factor of cut-set then a simple separation strategy suffices, for all SNR.

For examples of networks considered here, we refer the reader to Fig. 3.3 for a non-layered example or the one in Fig. 3.4 for a layered example.

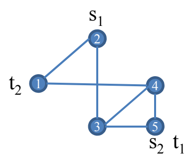
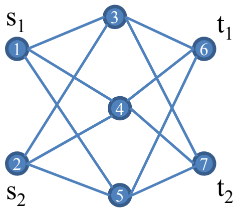


Figure 3.3: A multiple-unicast wireless network

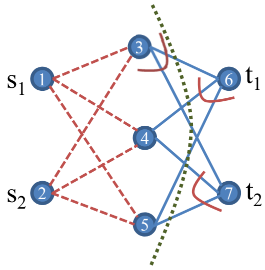
Theorem 20. *For a bidirected ergodic wireless network with k source destination pairs, the rate region given by*

$$\mathcal{R}^g(P) \supseteq \frac{\mathcal{R}^g_{cut}\left(\frac{P}{bd^3}\right)}{\mathcal{O}(\log k)} \tag{3.74}$$

is achievable using a separation strategy, where d is the maximum degree of any node and $b := \frac{e^{-\mathbb{E}(\log|h|^2)}}{2}$ is a constant depending on the fading distribution. If the



(a) Layered network



(b) Cuts

Figure 3.4: Layered network and cuts

fading is complex Gaussian, $b \approx 0.86$.

Proof. We will now show a scheme by which we can convert the bidirected ergodic wireless network into a bidirected polymatroidal network. We will view one snapshot of transmission in the network as a transmission in a bipartite graph, from the set of nodes $V^l = V$ on the left side to the set of nodes V^r on the right side. Node u on the left is connected to node v on the right with channel coefficient h_{uv} from the original network. The nodes have infinite memory, so each node is connected to itself with infinite capacity. We view the obtained network as a single-hop network and use a scheme for the single-hop (described in Sec. 3.4.5). The achievable rates for this single hop network are given by polymatroidal constraints, and therefore we obtain a polymatroidal network represented by a bi-partite graph with $2V$ nodes. This process can now be *reverted*, i.e., we can go from a layered representation with $2V$ nodes to a non-layered representation with V nodes. Thus we obtain a polymatroidal network with V nodes. There are edges E similar to the original graph, but the capacities are constrained according to the polymatroidal constraint:

$$\sum_{u \in \text{In}v} R_{uv} \leq \frac{1}{2} C(2P) \quad \forall v \in V \tag{3.75}$$

$$\sum_{u \in \text{Out}v} R_{vu} \leq \frac{1}{2} C(2P) \quad \forall u \in V. \tag{3.76}$$

This is now a bidirected polymatroidal network since the polymatroidal constraint is symmetric for the incoming and the outgoing edges at any given node. Now we perform routing over this bidirected polymatroidal network. Now the flow and cut on this network are related by:

$$\mathcal{R}_{\text{ach}}^{\text{g}} = \mathcal{R}_{\text{ach}}^{\text{poly}} \supseteq \frac{\mathcal{R}_{\text{cut}}^{\text{poly}}}{\mathcal{O}(\log k)}. \tag{3.77}$$

If we can relate the cut-set bound on the polymatroidal network and the cut-set bound on the original Gaussian network, we can get the desired result. This is done in the following lemma:

Lemma 29.

$$\mathcal{R}_{cut}^{poly}(P) \supseteq \frac{1}{2}\mathcal{R}_{cut}^{original}\left(\frac{P}{bd^3}\right). \tag{3.78}$$

Proof. See Appendix B.6 □

Now, (3.77) and (3.78) implies that

$$\mathcal{R}_{\text{ach}}^{\text{original}} \supseteq \frac{\mathcal{R}_{\text{cut}}^{\text{original}}\left(\frac{P}{bd^3}\right)}{\mathcal{O}(\log k)}, \tag{3.79}$$

which completes the proof of the theorem. □

Multi-antenna nodes: We can prove a result similar to the one in Theorem 20 even when there are multiple antennas at the nodes.

Theorem 21. *For a bidirected ergodic wireless network with multiple antenna nodes, the rate region for k-unicast is given by*

$$\mathcal{R}(P) \supseteq \frac{\mathcal{R}_{cut}\left(\frac{P}{bd^3}\right)}{\mathcal{O}(\log k)}, \tag{3.80}$$

is achievable using a separation strategy, where d is the maximum number of antennas that can communicate with any given antenna and $b := \frac{e^{-\mathbb{E}(\log|h|^2)}}{2}$ is a constant depending on the fading distribution.

Proof. The proof of the theorem is very similar to that of Theorem 20 except that when there are multiple antennas, each antenna is treated as a separate node with infinite capacity wireline links between antennas of the same node. This scheme can be shown to achieve the cut-set of this new network to within the approximation factor. We observe that any finite cut in this new network will partition the antennas in such a way that all antennas corresponding to the same original node will lie on the same side of the partition, since otherwise the value of the cut will become infinite. Therefore the cut-set in the new network and the original network have the same value and this completes the proof of the theorem. □

Special traffic scenarios: We now present results for *directed* fast fading networks under the special traffic patterns presented in Sec. 3.3.4. Since the network is directed, reciprocity will not be necessary to prove this result.

Theorem 22. *For a directed fast fading Gaussian network with multiple antenna nodes, a simple separation strategy can achieve a rate*

$$\mathcal{R}_{ach}(P) \supseteq \frac{\mathcal{R}_{cut}\left\{\frac{P}{bd^3}\right\}}{2} \quad \textit{for BC Traffic}, \tag{3.81}$$

$$R^{sum}_{ach}(P) \supseteq \frac{R^{sum}_{cut}\left\{\frac{P}{bd^3}\right\}}{2} \quad \textit{for X Traffic}, \tag{3.82}$$

$$R^{sum}_{ach}(P) \supseteq \frac{R^{sum}_{cut}\left\{\frac{P}{bd^3}\right\}}{4} \quad \textit{for group-communication traffic}, \tag{3.83}$$

where d is the maximum number of antennas that can communicate with any given antenna and $b := \frac{e^{-\mathbb{E}(\log|h|^2)}}{2}$ is a constant depending on the fading distribution.

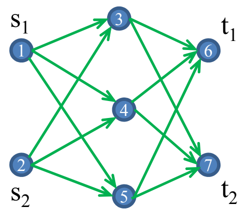
Proof. The proofs are similar to the proof of Theorem 20 and are therefore omitted. Here, the factor loss of 2 is present for the BC scenario due the factor 2 loss in the local physical layer scheme (see Theorem 13). □

Special channel model: directed layered network

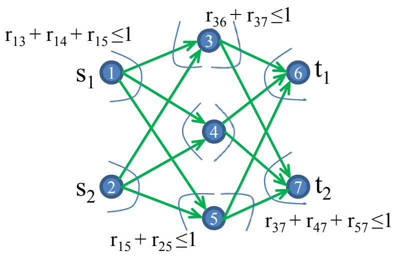
In this section, we consider directed fully-connected (f.c.) layered networks. These are layered networks, which have connectivity between adjacent layers only in the forward direction, i.e., links are always between a node in V_i to a node in V_{i+1}. Further for a fully-connected network, we assume that $(u, v) \in \mathcal{E} \quad \forall u \in V_i, v \in V_{i+1}$. Consider, for example, the network in Fig. 3.5a.

Theorem 23. *For a directed fully-connected layered ergodic wireless network with k-distinct sources in the first layer having messages to k distinct sinks in the last layer, the rate region given by*

$$\mathcal{R}(P) \supseteq \frac{\mathcal{R}_{cut}\left\{\frac{P}{bd^3}\right\}}{2} \tag{3.84}$$



(a) Wireless network



(b) Polymatroidal network

Figure 3.5: Directed layered F.C. networks

is achievable using a separation strategy, where d is the maximum degree of any node and $b := \frac{e^{-\mathbb{E}(\log|h|^2)}}{2}$.

Proof. The achievable scheme is again a local physical layer scheme, i.e., for each hop in the layered network we use the strategy for X-channels described in Sec. 3.4.5. Now we get a special type of directed polymatroidal network for which we can show that max-flow equals min-cut.

Theorem 24. *For a k-unicast problem in a polymatroidal directed layered network in which all the node constraints are of the form*

$$\sum_{u \in In\{v\}} R_{uv} \leq 1, \quad \forall v, \tag{3.85}$$

$$\sum_{v \in Out\{u\}} R_{uv} \leq 1, \quad \forall u, \tag{3.86}$$

the rate region given by max-flow equals the min-cut for the k-unicast problem.

Proof. See Appendix B.7. □

Now we can use Lemma 29 to get:

$$\mathcal{R}_{\text{ach}}^{\text{original}}(P) = \mathcal{R}_{\text{ach}}^{\text{poly}}(P) = \mathcal{R}_{\text{cut}}^{\text{poly}}(P) \supseteq \frac{1}{2}\mathcal{R}_{\text{cut}}^{\text{original}}\left\{\frac{P}{bd^3}\right\}, \tag{3.87}$$

which implies the theorem. □

3.5.4 Fixed Gaussian network

Consider a general Gaussian network where broadcast and superposition can occur simultaneously, and where the channel coefficients are fixed but drawn from a continuous distribution. We studied a simple single-hop version of such a network in Sec. 3.4.6. We will now show that the local scheme can be placed in a network context to get an almost-sure DOF characterization.

Theorem 25. *For a bidirected fixed multi-antenna Gaussian network with k source destination pairs, the DOF given by*

$$\mathcal{D}_{ach} \supseteq \frac{\mathcal{D}_{cut}}{\mathcal{O}(\log k)} \tag{3.88}$$

is achievable almost surely *using a separation strategy.*

Proof. The proof proceeds analogously to Theorem 21, but instead of using the local scheme in Theorem 13, the scheme of Theorem 14 is used and therefore a DOF characterization is obtained. □

Special traffic scenarios: We now present results for *directed* Gaussian networks under the special traffic patterns presented in Sec. 3.3.4. Since the network is directed, reciprocity will not be necessary to prove this result.

Theorem 26. *For a directed Gaussian network with each channel coefficient chosen from a continuous distribution, a simple separation strategy can achieve a DOF region,*

$$\begin{aligned} \mathcal{D}_{ach} &\supseteq \frac{\mathcal{D}_{cut}}{2} \quad \textit{for BC Traffic}, && (3.89)\\ D_{ach}^{sum} &\supseteq \frac{D_{cut}^{sum}}{2} \quad \textit{for X Traffic}, && (3.90)\\ D_{ach}^{sum} &\supseteq \frac{D_{cut}^{sum}}{4} \quad \textit{for group-communication traffic}. && (3.91) \end{aligned}$$

3.5.5 Linear deterministic networks with MAC and broadcast components

A linear deterministic network composed of MAC and broadcast components is defined in the same way as the Gaussian network composed of MAC and broadcast components. The key difference is that the transmissions are over a finite field $\mathbb{F}_q$, there is no noise, and the channels between node i and j of color c are matrices in general: H_{ij}^c. The main result for linear deterministic networks with MAC and broadcast components is stated in the following theorem:

Theorem 27. *For the k-unicast problem in linear deterministic network composed of broadcast and multiple access channels, the layered architecture can achieve a rate,*

$$\mathcal{R}_{ach}^{ld} \supseteq \frac{\mathcal{R}_{cut}^{ld}}{\mathcal{O}(\log k)}. \tag{3.92}$$

Proof. The proof of this theorem proceeds similarly to that of Theorem 16, the key difference being the fact that for linear deterministic networks, the cut-set bounds under product form and general distributions are the same, and therefore, there is no power scaling factor. $\square$

Special traffic scenarios: We now present results for *directed* linear deterministic networks with MAC and broadcast components under the special traffic patterns presented in Sec. 3.3.4.

Theorem 28. *For a directed linear deterministic network composed of broadcast and multiple access channels, a simple separation strategy can achieve a rate*

$$\begin{aligned}
\mathcal{R}_{ach} &\supseteq \mathcal{R}_{cut} \quad \textit{for BC Traffic}, && (3.93)\\
R_{ach}^{sum} &\supseteq R_{cut}^{sum} \quad \textit{for X Traffic}, && (3.94)\\
R_{ach}^{sum} &\supseteq \frac{R_{cut}^{sum}}{2} \quad \textit{for group-communication traffic}. && (3.95)
\end{aligned}$$

Proof. The proofs are similar to the proof of Theorem 27 and are therefore omitted. $\square$

3.5.6 Networks of fast fading MAC and broadcast channels with delayed CSIT

We will now consider networks composed of fast fading MAC and Broadcast channels where the channel states are i.i.d. over antennas and time, and the CSI is available at the transmitting nodes after a delay. Schemes for each of these local channels were studied in Sec. 3.4.4 and a degree of freedom characterization was obtained. Now, we will try to obtain the degree of freedom region for multiple unicast over a network of such channels.

Our main result for such networks is the following.

Theorem 29. *For the k unicast problem in Gaussian network composed of fading broadcast and multiple access channels with delayed feedback, a simple separation strategy can achieve a DOF,*

$$\mathcal{D}_{ach} \supseteq \frac{\mathcal{D}_{cut}}{\mathcal{O}(\log k)\mathcal{O}(\log p_{\max})}, \tag{3.96}$$

where

$$p_{\max} = \max_{BC\ channels} p(BC\ channel), \tag{3.97}$$

and $p(BC\ channel)$ is given by the minimum of number of transmit antennas and the total number of received antennas in the broadcast channel.

Proof. The coding scheme is again a separation-based strategy: each component broadcast or multiple access channel with delayed feedback is coded for independently creating bit-pipes on which information is routed globally. The physical layer technique is the scheme for MIMO broadcast and MAC channels with delayed CSIT proposed in Sec. 3.4.4. The proof is very similar to that in Sec. 3.5.2.

For the fading MIMO broadcast channel with delayed feedback, we can achieve the following DOF region $\mathcal{D}_{\text{ach}}^{\text{BC}}$ with feedback. For the fading MIMO multiple access channel, we can achieve the region given by $\mathcal{D}_{\text{ach}}^{\text{MAC}}$, but we will restrict ourselves to achieve the smaller rate region $\mathcal{D}_{\text{ach}}^{\text{BC}}$ for the purpose of symmetry. This induces a bidirected polymatroidal network, in which we can use flows to achieve a rate region $R_{\text{ach}}^{\text{poly}}$. By Theorem 12, we have

$$R_{\text{ach}}^{\text{poly}} \supseteq \frac{\mathcal{R}_{\text{cut}}^{\text{poly}}}{\mathcal{O}(\log k)}. \tag{3.98}$$

By Lemma 24, we have $\mathcal{D}_{\text{ach}}^{\text{BC}} \supseteq \frac{\mathcal{D}_{\text{cut}}^{\text{BC}}}{\mathcal{O}(\log p_{\max})}$ and by choice, $\mathcal{D}_{\text{ach}}^{\text{MAC}} = \frac{\mathcal{D}_{\text{cut}}^{\text{MAC}}}{\mathcal{O}(\log p_{\max})}$. Further, since cuts in the polymatroidal and the original fading network decompose into cuts for each channel, any cut-set in the polymatroidal network induced by the achievable scheme has a counterpart cut-set in the erasure network within a factor of $\mathcal{O}(\log p_{\max})$:

$$\mathcal{R}_{\text{cut}}^{\text{poly}} = \frac{\mathcal{D}_{\text{cut}}^{\text{fading}}}{\mathcal{O}(\log p_{\max})}. \tag{3.99}$$

Also, we can achieve the same DOF in the original fading network as the rate in the polymatroidal network, i.e.,

$$\mathcal{R}_{\text{ach}}^{\text{poly}} = \mathcal{D}_{\text{ach}}^{\text{fading}}. \tag{3.100}$$

This implies that

$$\mathcal{D}_{\text{ach}}^{\text{fading}} \supseteq \frac{\mathcal{D}_{\text{cut}}^{\text{fading}}}{\mathcal{O}(\log^3 k)\mathcal{O}(\log p_{\max})}. \tag{3.101}$$

This completes the proof of the theorem. □

Special traffic scenarios: We now present results for *directed* networks with MIMO broadcast and MAC channels with delayed feedback under the special traffic patterns presented in Sec. 3.3.4. Since the networks are directed, reciprocity is not needed; however, we will continue to assume that the broadcast channel gets delayed CSI feedback.

Theorem 30. *For a directed network composed of MIMO broadcast and MAC channels with delayed CSI feedback, a simple separation strategy can achieve a DOF region*

$$\begin{aligned}
\mathcal{D}_{ach} &\supseteq \frac{\mathcal{D}_{cut}}{\log(p_{\max}+1)} && \textit{for BC Traffic}, && (3.102)\\
D_{ach}^{sum} &\supseteq \frac{D_{cut}^{sum}}{\log(p_{\max}+1)} && \textit{for X Traffic}, && (3.103)\\
D_{ach}^{sum} &\supseteq \frac{D_{cut}^{sum}}{2\log(p_{\max}+1)} && \textit{for group-communication traffic}, && (3.104)
\end{aligned}$$

where

$$p_{\max} = \max_{\textit{BC channels}} p(\textit{BC channel}), \tag{3.105}$$

and $p(\textit{BC channel})$ is given by the minimum of number of transmit antennas and the total number of received antennas in the broadcast channel.

Proof. The proofs are similar to the proof of Theorem 29 and are therefore omitted. □

3.5.7 Fading linear deterministic network

Consider a linear deterministic network of the form defined in Sec. 3.4.7, where each of the non-zero channel coefficients $H_{ij}(t)$ undergoes i.i.d. fading with a uniform distribution on the non-zero elements. Then the local scheme of Sec. 3.4.7 can be extended to a global network scheme.

Theorem 31. *For a bidirected linear deterministic network with k source destination pairs, the rate region given by*

$$\mathcal{R}_{ach} \supseteq \frac{\mathcal{R}_{cut}}{\mathcal{O}(\log k)} \tag{3.106}$$

is achievable using a separation strategy.

Proof. The proof proceeds analogously to Theorem 20, but instead of using the local scheme in Theorem 13, the scheme of Theorem 15 is used. □

Special traffic scenarios: We now present results for *directed* fast fading linear deterministic networks under the special traffic patterns presented in Sec. 3.3.4. Since the network is directed, reciprocity will not be necessary to prove this result.

Theorem 32. *For a directed fast fading linear deterministic network, a simple separation strategy can achieve a rate*

$$\begin{aligned} \mathcal{R}_{ach} &\supseteq \frac{\mathcal{R}_{cut}}{2} \quad \textit{for BC Traffic}, && (3.107) \\ R_{ach}^{sum} &\supseteq \frac{R_{cut}^{sum}}{2} \quad \textit{for X Traffic}, && (3.108) \\ R_{ach}^{sum} &\supseteq \frac{R_{cut}^{sum}}{4} \quad \textit{for group-communication traffic}. && (3.109) \end{aligned}$$

CHAPTER 4

FUNCTION COMPUTATION PROBLEM

"Sometimes, a question is the answer too!" – Akshara

4.1 Function Computation

Having solved the multiple-unicast problem in wireless networks, we now turn our attention to the more general problem of function computation. In several communication scenarios, a receiver is interested in computing a function of data from different agents spread over a network. For example, in a sensor network, a fusion node is interested in computing a function of the various sensors. Similarly, in a cloud computing scenario, the data may be stored in a distributed manner, with different types of information about the same record stored in different locations. In such a scenario, one may be interested in computing a function of the data.

The function computation problem in undirected graphs is defined as follows: there are K independent function-computation sessions, each involving the computation of a function of S independent sources at a specified terminal. While the data involved in distinct sessions are assumed to be independent, all the communication happens via a common communication infrastructure, modeled as a capacitated undirected graph $G = (V, E)$ and capacity function functions $c(e), e \in E$.

This multi-session function computation scenario has several practical applications: in a sensor network, a communication infrastructure may be shared by several sensors of different modalities like heat, pressure, infrared and each of these modalities has a distinct or common fusion center.

For an instance of this problem, see Fig. 4.1, where there are 2 sessions. In the first session, sensing nodes a and d have temperature information X_1 and X_2, and node b, which is the fusion node for temperature sensing, wants to compute the average temperature $\frac{X_1+X_2}{2}$. In the second session, sensing nodes b, e, d have pressure information Y_1, Y_2, Y_3, and node c, which is the fusion node for pressure

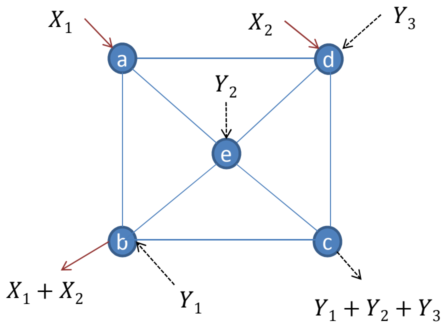


Figure 4.1: Multi-session function computation example

sensors, wants to compute the average pressure $\frac{Y_1+Y_2+Y_3}{3}$.

We consider the setting of block function computation, where receiver i is interested in computing R_iT function evaluations at the end of T time instants (the nodes in the network can forward arbitrary functions of data that they receive). The performance of the scheme is measured by the set of rates $(R_1, ..., R_K)$, which can be achieved, this is called the computation capacity region $\mathcal{C}$.

A natural achievable strategy in this context is the use of *computation trees*, by which we mean strategies where a tree spanning the vertices involved in function computation is constructed (called a Steiner tree), and along the tree, in-network computations are performed in such a way that the function can be computed at the destination. Note that this is a very simple in-network computation strategy and does not involve block coding of data. Furthermore, this strategy avoids inter-session network coding, i.e., mixing of information across the different sessions (or modalities). These features make it a feasible practical scheme. We will show the surprising result that such a simple strategy can achieve near optimal performance.

The proposed achievable strategy is *universal*, i.e., it works for the computation of an arbitrary function. However, the performance of the strategy depends on a property of the function called λ-divisibility. A function f is said to be λ_f-*divisible* if the function can be computed in a divide-and-conquer fashion with every intermediate computation requiring at most λ_f symbols to store and communicate. In other words, every Steiner tree can be used for computation of the function, i.e., every Steiner tree is a computation tree with computation rate $\frac{1}{\lambda_f}$. Every function on S variables is λ_f-divisible for some $\lambda_f \leq |S|$. For efficient computation of the function, we would like λ_f to be as small as possible.

It turns out that we can write a cut-set upper bound for computation of a class of functions that satisfy the marginal-injectivity property. A function is said to be *marginally injective* if it is injective with respective to every variable, given any realization of the remaining variables. Note that, for a marginally injective function f, $\lambda_f \geq 1$. For example, linear functions over fields or commutative groups are marginally injective, and have $\lambda_f = 1$.

In this thesis, we show that for function computation in a graph $\mathcal{G}$ with K sessions and S non-overlapping sources per session, computation trees achieve $\mathcal{R}_{\text{comp}}$ such that

$$\frac{\bar{C}}{\lambda_f(S)g(S,K)} \subseteq \mathcal{R}_{\text{comp}} \subseteq \bar{C}, \tag{4.1}$$

where $\bar{C}$ is the cut-set bound, where

$$g(S,K) \leq \begin{cases} 1 & \text{if } \mathcal{G} \text{ is a tree} \\ 2 & \text{if } K = 1, \\ O(\log SK) & \text{if } K > 1. \end{cases}$$

Furthermore, there are polynomial-time algorithms to find the computation trees that achieve these performance guarantees.

This result is proved by connecting the function computation problem to the problem of approximating "sparsest Steiner cuts" by using Steiner flows, for which similar results are known [77, 108, 56]. Our result incorporates several special cases: when $S = 1$, it is the multiple unicast problem with K source-destination pairs, for which the seminal work of Leighton and Rao [90] shows a $O(\log K)$ gap between routing and cuts; when $K = 1$, it is the function computation scenario with a single receiver being interested in a function of S nodes. For all of these cases, we demonstrate that there is a close connection between the function computation problem and the multiple multicast communication problem, where there are K independent sources each of which wants to send common messages to a set of S destinations.

This result demonstrates that undirected graphs are fundamentally different from directed graphs in various aspects. While directed graphs with cycles are more general than undirected graphs, the special structure imposed by directed graphs allows very efficient packings of Steiner trees, thereby achieving performance close to the cut. For general directed graphs, Steiner tree packing can achieve a performance arbitrarily far away from the cut [3]. Furthermore, there are no known algorithms to compute the best Steiner tree packings within a constant factor approximation in directed graphs [23], whereas there is a polynomial time algorithm to compute a factor-2 approximation for Steiner tree packings in undirected graphs [154]. It should be noted that the presence of cycles in undirected graphs allows for interactive function computation. Nevertheless, our results show that, for several function classes, simple non-interactive function computation is near optimal.

4.1.1 Related work

Communication complexity: The problem of function computation has been well studied in the communication complexity literature. The basic setup is that there are two nodes, having values x and y, and both of them wish to compute a function $f(x, y)$ of these nodes. The goal is to minimize the total number of bits communicated in the worst case. This problem was originally formulated by [158] and the reader is referred to [85] for a treatment of this problem. However, this setting does not allow for a block computation of the function.

Function computation in random graphs: The problem of block function computation was originally formulated by Giridhar and Kumar [51]. They studied the scaling laws for computing several classes of functions in a random wireless network. In their work, the wireless nature of the medium is dealt with using scheduling (to avoid interference) and then in-network computation is performed by aggregating data through spanning trees. In contrast, in this thesis, we do not deal with the wireless aspects and consider the problem of function computation on any specified undirected graph.

Computation trees: Computation trees, where Steiner trees are used for computation, have been a popular strategy (see [51, 82, 12]), although the name was coined later in [133]. Such trees are optimal for function computation when the graph is itself just a directed tree [82], [12]. In an arbitrary directed acyclic graph, [12] proposed a strategy where several such trees can be packed and the rate of computation can be linked to the Steiner packing number. However, as pointed out earlier, in directed graphs, the Steiner packing number can be arbitrarily far away from the cut [3].

In undirected graphs, packing of specific computation trees resembling the structure of the function was considered, and algorithms for packing computation trees were provided in [133]. However, there are no guarantees on the computation rates achieved by this algorithm and it is unknown what the gap to capacity is.

Linear coding: Another popular achievable strategy is linear coding where each intermediate node forwards a linear combination of the incoming symbols. For general directed graphs, when the function to be computed is linear, random linear coding is known to be optimal [127]. In [13], it is shown that linear coding is insufficient when the function to be computed is nonlinear and the potential loss due to employing linear coding is quantified.

Undirected graphs: The problem of determining function computation capacity in undirected graphs is considered harder in general, due to the presence of cycles in the graph, which allow for *interaction* in function computation. Single-shot function computation in the 2-node setting has been a central problem of study in the field of communication complexity [158, 85]. Even allowing for block computation, for general functions, seemingly simple 2-node problems can become hard (see [5]). For a class of sum-threshold functions on undirected trees, this capacity is characterized in [82] using carefully orchestrated interactive strategies. In this thesis, we show that for many function classes of practical interest, non-interactive function computation using computation trees can give near-optimal performance.

Multi-session function computation: The case of function computation with multiple sessions has not received much attention, partly because even the single terminal scenario is sufficiently complicated and unsolved in general directed graphs. One particular problem that has attracted attention is the case where there are several sources in a directed acyclic graph, and each destination demands the sum of the sources [128, 87, 126]. This setup is different from the setting in the current paper, where each destination demands *general functions* of *distinct variables*.

Correlated sources: Information theoretic approaches to function computation in simple settings with correlated sources were studied by [95, 99, 34]. Function computation through noisy channels was studied in [161, 35, 18] and function computation through Gaussian wireless channels was studied in [111, 117]. However, these results are restricted to simple functions. In this thesis, we only consider independent sources and noiseless channels but focus on general graphs and a wide class of functions.

4.1.2 Organization

We will first discuss the mathematical formulation of the problem and describe the classes of functions considered in the paper with examples in Section 4.2. Next, we describe our proposed scheme and outer bound in Section 4.3. We state our main result and discuss its ramifications, especially the connection between the computation problem and a dual communication problem, in Section 4.4. Finally, we prove the main result in Section 4.5.

4.2 Problem Formulation

The communication network is represented by an undirected graph $\mathcal{G} = (V, E)$, and edge capacity function $c : E \to \mathbb{R}^+$. There are K "sessions," each involving independent variables and the computation of a specified function at a terminal. The sources for session k are denoted by S_k, where

$$S_k = \{\sigma_{1k}, \sigma_{2k} ..., \sigma_{Sk}\} \subseteq V, \quad k = 1, 2, ..., K, \tag{4.2}$$

each with its corresponding destination $\rho_k, k = 1, 2, .., K$. Define $G_k := S_k \cup \{\rho_k\}$ as the group of vertices involved in session k. We assume for notational simplicity that each session involves the same number of variables; in the general case that session k involves $|S_k|$ variables, we can replace SK in the results by $\sum_k |S_k|$ and the results will continue to hold.

The source σ_{ik} has several instances of an information variable X_{ik}, which takes values over a common alphabet $\mathcal{A}$. We do not impose a statistical structure on the messages. Alternatively, it is possible to set up the sources as independent random variables, in which case, the results in the paper will continue to hold with modification of notation. The destination ρ_k is interested in computing a function $f : \mathcal{A}^S \to \mathcal{B}$, i.e., it wishes to reconstruct $f(X_{1k}, X_{2k} .., X_{Sk})$. While S_k and $S_{k'}$ could in general overlap, we assume that the information of interest to destination ρ_k is distinct from the information of interest to destination $\rho_{k'}$.

This function computation happens several times and therefore we consider

Table 4.1: Function Examples

	Function Name	$f(x_1, ..., x_S)$	$\lambda_f(S)$-divisible	Marginally Injective?
1	Linear (field)	$x_1 \oplus x_2 \oplus ...x_S$	1	✓
2	Addition (Abelian group)	$x_1 \oplus x_2 \oplus ...x_S$	1	✓
3	Arithmetic Sum ($A := \|\mathcal{A}\|$)	$x_1 + ... + x_S$	$\log_A\{S(A-1)+1\}$	✓
4	Real Sum over $\mathcal{A} = \{0 : \delta : 1\}$	$x_1 + ... + x_S$	$\log_A\{S(A-1)+1\}$	✓
5	ℓ_p-norm suitably quantized	$x_1^p + ... + x_S^p$	$\log_A\{S(A-1)+1\}$	✓
6	Histogram	Hist($x_1, ..., x_S$)	$\log_A\{\binom{S+A-1}{S}\}$	✓
7	Symmetric	f($x_1, ..., x_S$)	$\leq \log_A\{\binom{S+A-1}{S}\}$	Depends
8	Maximum	$\max(x_1, ..., x_S)$	1	No
9	Sum threshold	$\mathbb{1}_{[x_1+x_2+...+x_S>m]}$	$\log_A\{S(A-1)+1\}$	No

a block function computation problem as follows. The network can potentially employ in-network computation (network coding) in order to compute the function. Formally, a coding scheme of block length T and achieving a rate tuple $(R_1, ..., R_k)$ is described as follows:

- The source σ_{ik} has messages $X_{ik}(1), X_{ik}(2), ..., X_{ik}(R_k T)$. The destination ρ_k is interested in computing $f(X_{1k}(t), X_{2k}(t), ..., X_{Sk}(t))$ for $t = 1, 2, ..., R_k T$. The message set of source $v = \sigma_{ik}$ is denoted by W_{vk} where $|W_{vk}| = |\mathcal{A}|^{R_k T}$. For simplicity of notation, if a given node $v \notin S_k$, then we set $W_{vk} = \emptyset$.

- Since the network graph is undirected the capacity $c(e)$ on edge e has to be shared between the forward and reverse direction, let the fraction on forward direction be $\alpha(e)$. Once the α is fixed, then the network becomes a directed network. Let us scale this network to have integral capacities and represent this network as a *directed multigraph*, potentially having multiple edges between the same nodes.

- At each node v, at each time t, there is a mapping

$$g_{v,t} : \mathcal{A}^{|\text{In}(v)|(t-1)} \times W_{v1} \times W_{vk} \times ...W_{vK} \rightarrow \mathcal{A}^{|\text{Out}(v)|}, \tag{4.3}$$

  where In(v) and Out(v) denote the set of incoming edges and the set of outgoing edges of node v. The function $g_{v,t}$ thus specifies a mapping from the messages on the incoming edges till time $t-1$ and the node's own messages to the outgoing edge message at time t.

- The decoding map at destination ρ_k is given as a function of its incoming edges till time t

$$\psi : \mathcal{A}^{|\text{In}(v)|T} \rightarrow \mathcal{B}^{R_k T}. \tag{4.4}$$

- The coding scheme is said to achieve rate tuple $(R_1, ..., R_K)$ if each destination correctly recovers the function of its desired nodes.

The set of all achievable rate tuples $(R_1, ..., R_K)$ is called the computation capacity region $\mathcal{C}$.

4.2.1 Function classes

We will define certain classes of functions which we will be interested in. A function $f : \mathcal{A}^S \to \mathcal{B}$ is called λ_f-divisible, if for every index set $I \subseteq [S]$, there exists a finite set $\mathcal{B}_I$ and a function $f^I : \mathcal{A}^{|I|} \to \mathcal{B}_I$ such that the following hold:

1. $f^{[S]} = f$.
2. $|f^I(.)| \leq |\mathcal{A}|^{\lambda_f}$.
3. For every partition $\{I_1, ..., I_j\}$ of I, there exists a function $g : \mathcal{B}_{I_1} \times \cdots \times \mathcal{B}_{I_j} \to \mathcal{B}_I$ such that for every $x \in \mathcal{A}^|I|$,

$$f^I(x) = g(f^{I_1}(x_{I_1}), ..., f^{I_j}(x_{I_j})). \tag{4.5}$$

If a function is λ_f-divisible, then it means that it can be computed in a divide-and-conquer manner such that every intermediate computation needs only λ_f symbols to store and transmit. Every function is $|S|$-divisible in a trivial manner, since retaining all the information is sufficient to compute the function. Our interest will be in functions f for which λ_f is small. λ_f-divisible functions can be seen to be equivalent to λ_f-bounded functions defined in [12]; we prefer the alternate name and definition since it is more suggestive. Divisible functions as defined in [51, 138] and [12] are λ_f-divisible with $\lambda_f = \log_{|\mathcal{A}|} |\mathcal{B}|$. For example, a linear function over a finite field has $\lambda_f = 1$. We refer the reader to Table 4.1 for a listing of λ_f for various functions.

A function $f : \mathcal{A}^S \to \mathcal{B}$ is called *marginally injective* if, for any variable i, for any fixed assignment on the other variables, the function can take on $\mathcal{A}$ distinct values as x_i varies over $\mathcal{A}$, i.e.,

$$\forall i, \forall y_{[S]\setminus i} \in \mathcal{A}^{S-1}, \psi : \mathcal{A} \to \mathcal{B}, \tag{4.6}$$

defined by $\psi(x_i) = f(x_i, y_{[S]\setminus i})$, is injective.

Several functions of interest satisfy this property, as listed in Table 4.1. An example of a function which is *not* marginally injective is the max-function over an ordered set. The value of the maximum does not depend on any other variable if one of the variables is assigned the maximum possible value.

4.2.2 Examples

Various examples of functions are provided in Table 4.1. Also listed is whether the function is marginally injective or not, and the value of λ_f for which f is λ_f-divisible (the value of λ_f could in general depend on the number of variables S).

1. The linear function over a finite field is an obvious example of a function which is 1-divisible and marginally injective. For linear function computation with a single terminal, linear coding is known to be optimal [127], but the case of multiple terminals has not been studied.

2. The case of addition over an Abelian group is very similar to the finite field sum. However, for addition over an Abelian group, existing random linear coding techniques do not apply due to the lack of the field structure, whereas our computation tree based approach naturally extends to this case.

3. The arithmetic sum is $\log_A B$-divisible since maintaining the arithmetic sum of the subsets is sufficient, and it is also marginally injective.

4. Real sum over a quantized alphabet on [0,1] (quantized to a fixed precision δ) is only a disguised version of the arithmetic sum since after scaling by $\frac{1}{\delta}$, we have converted it into the arithmetic sum. If precise quantization is not required in the application, then we can maintain quantized versions of the sum (to accuracy δ) throughout the computation tree, thus having an effective λ of 1. This function is practically relevant in sensor networks, since sum of the LLR (log-likelihood-ratio) is a sufficient statistic in some cases [10].

5. ℓ_p-norm is basically just a real sum, except for the fact that we need x_i^p to be quantized to a precision of δ. This is again a practically important function in sensor networks.

6. The histogram is a function from $\mathcal{A}^S$ to $\mathcal{B}$ with $A := |\mathcal{A}|$ and $B := |\mathcal{B}|$. Using a simple enumeration it can be computed that the histogram can take on one of $B = \binom{S+A-1}{S}$ values. The histogram is $\log_A B$-divisible because, for any subset, the histogram of the subset is a sufficient statistic to maintain. For the case of computing a histogram of S nodes at a single terminal, our method leads to a $2\lambda_f = 2\log_A \binom{S+A-1}{S}$ approximation as opposed to the

method in [12] of using the binary arithmetic sum, which leads to a bigger factor gap of $(A-1)\log_A(AS)$ (adapted to undirected graphs). For example, when $A = 16$, $S = 50$, $2\lambda_f = 23.6$, whereas $(A-1)\log_A(AS) = 36.1$.

7. Any symmetric function (which is invariant to permutations of input symbols) depends only on the histogram. Thus the histogram is a sufficient statistic to compute any symmetric function.

8. Maximum is an example of a function which is clearly 1-divisible, however it is not marginally injective. Consider the alphabet $0, 1, 2, .., a-1$. If we know that $\max(x_2, ..., x_S) = A - 1$, then the function f no longer depends on x_1 and is therefore not marginally injective. While the achievable strategies in this thesis continue to hold, the outer-bound is no longer valid for this function.

9. Sum-threshold functions are basically of the form $\mathbb{1}_{[x_1+x_2+...x_S>m]}$. The arithmetic sum of the variables is a sufficient statistic to compute the function and hence the function has

$$\lambda_f \leq \log_A\{S(A-1)+1\}. \tag{4.7}$$

The function is not marginally injective, because, if the variables other than x_i have a sum of greater than m, the function no longer depends on x_i.

4.3 Inner and Outer Bounds

In this section, we present our achievable scheme based on computation trees and the outer-bound on the rates for multi-session function computation in undirected graphs. Similar bounds have been previously studied in the context of single-session function computation in directed acyclic graphs [12].

4.3.1 Cut-set bound

The cut-set bound defined in [12] is special to directed acyclic graphs and does not generalize to cyclic graphs. We use the technical condition of *marginal injectivity* in order to establish a simple cut bound on the communication rate. Given a set

$\Omega \subseteq V$, define

$$K(\Omega) := \{k : G_k \cap \Omega \neq \emptyset, G_k \cap \Omega^c \neq \emptyset\} \tag{4.8}$$

as the set of sessions disconnected by Ω and define

$$\text{Cut}(\Omega) := \sum_{(ij)\in E: i\in\Omega, j\in\Omega^c} c(ij). \tag{4.9}$$

Then, for any scheme computing a *marginally injective* function, for any $\Omega \subseteq V$, define

$$\bar{\mathcal{C}} = \{\overline{R} : \sum_{k\in K(\Omega)} R_k \leq \text{Cut}(\Omega)\ \forall\Omega\}. \tag{4.10}$$

The main observation is that $\mathcal{C} \subseteq \bar{\mathcal{C}}$. We will prove this for the case of $K = 1$, and the general case is similar. When $K = 1$, we only need to consider Ω such that $K(\Omega) = 1$, i.e., the cut separates G_k. Let Ω be such that $\rho_1 \in \Omega^c$ and for some i, $\sigma_{i1} \in \Omega$. The information on edges between Ω and Ω^c can take $\mathcal{A}^{\text{Cut}(\Omega)T}$ possible values. The function is marginally injective, and therefore this should at least convey as much information as one of the sources, σ_{1i}. Thus

$$|\mathcal{A}|^{\text{Cut}(\Omega)T} \geq |\mathcal{A}|^{R_1 T}, \tag{4.11}$$

and so $R_1 \leq \text{Cut}(\Omega)$ which proves the required bound.

Informally, the marginal-injectivity property suggests that even in the presence of feedback or interaction, each node needs to convey its information symbol in order for function computation to succeed. Therefore, we can think of marginally-injective functions as functions for which interaction does not help much. On the other hand, if a function is not marginally injective, interaction can in general help.

4.3.2 Achievable strategy

We first formally define Steiner trees. An *undirected Steiner tree* on a set G is defined as an undirected tree τ which includes G in its vertex set. A *directed Steiner tree* rooted at ρ on a vertex set S is defined as a directed tree rooted at ρ that includes S in its vertex set. Given an undirected Steiner tree on $G_k := \{\rho_k\} \cup S_k$,

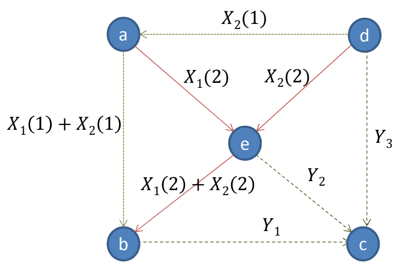


Figure 4.2: Multi-session example

there is a (unique) orientation of edges such that we get a directed Steiner tree on S_k rooted at ρ_k. Note that whenever we mention about Steiner tree, we refer to a tree with unit capacity edges.

Let $\mathcal{T}_k$ be the set of Steiner trees on G_k and $\mathcal{T} = \cup_k \mathcal{T}_k$. A fractional Steiner packing of $\pi_1, ..., \pi_K$ is said to be achievable if so many trees can be packed simultaneously: i.e., there exist $f_\tau, \tau \in \mathcal{T}$ such that

$$\pi_k = \sum_{\tau \in \mathcal{T}_k} f_\tau \quad k \in [K] \tag{4.12}$$

$$\sum_{\tau \in \mathcal{T}: e \in \tau} f_\tau \leq c(e) \quad \forall e \in E. \tag{4.13}$$

Let $\mathcal{R}_s$ be the Steiner packing rate region, i.e., the set of all fractional Steiner packing rates $(\pi_1, ..., \pi_K)$, which are achievable.

The achievable strategy is based on using computation trees. The key observation is that, if a function is λ_f-divisible, then any Steiner tree with unit capacity can be used to compute the function at a rate $\frac{1}{\lambda_f}$. Thus a rate tuple

$$(R_1, ..., R_K) = \frac{1}{\lambda_f}(\pi_1, \pi_2, ..., \pi_K) \tag{4.14}$$

can be achieved for computing any λ_f*-divisible function*. We denote the set of all rate tuples achievable by this computation tree scheme as $\mathcal{R}_{\text{comp}}$. Thus

$$\mathcal{R}_{\text{comp}} = \frac{\mathcal{R}_s}{\lambda_f}. \tag{4.15}$$

For a demonstration of our achievable strategy for the example function computation problem in Fig. 4.1, see Fig. 4.2. In this strategy, there are two Steiner trees for session 1, the first comprised of edges da and ab, and the second is comprised of edges ae, de and eb. There is one Steiner tree for session 2 which has edges bc, ec and dc. This Steiner packing strategy achieves the sum rate. This can be observed by taking the cut separating b from the rest of the nodes; since the cut separates both sessions the corresponding value of the cut is a bound on the sum rate. Note that in general, in our achievable strategy, the capacity of each edge could be shared between several Steiner trees; this is called fractional Steiner packing, as opposed to the integral Steiner packing demonstrated in this example.

We note that both our achievable rate and outer bound depend only on the definition of the sets G_k and not on the particular way in which $G_k = S_k \cup \{\rho_k\}$ is divided into sources S_k and the destination ρ_k. Thus if the destination ρ_k swaps its role with one of the sources in S_k, the achievable rate and the outer bound remains the same. It is not clear if this property holds for any achievable strategy and outer bound; in particular, it is interesting to study this question for the capacity region.

4.4 Main Result

We first state our main result, which shows a factor approximation for the capacity region.

Theorem 33. *For computation of λ-divisible functions in a graph $\mathcal{G}$ with SK sources and K terminals demanding functions of S non-overlapping variables, computation trees achieve $\mathcal{R}_{comp}$ such that*

$$\frac{\bar{\mathcal{C}}}{\lambda_f(S)g(S,K)} \subseteq \mathcal{R}_{comp}, \tag{4.16}$$

where $\bar{C}$ is the cut-set bound, and

$$g(S,K) \leq \begin{cases} 1 & \textit{if } \mathcal{G} \textit{ is a tree} \\ 2 & \textit{if } K = 1, \\ O(\log SK) & \textit{if } K > 1. \end{cases}$$

is a function that does not depend on the number of nodes in the network or the specific function to be computed. Furthermore, if the function is marginally injective, $\mathcal{C} \subseteq \bar{\mathcal{C}}$.

In the most general case, this result shows that our achievable strategy is optimal to within a factor $\lambda_f(S)O(\log SK)$. For functions which have a small λ_f, (for example, linear functions which have $\lambda_f(S) = 1$), the dominating factor in the approximation is $O(\log SK)$. We will compare this to some simple bounds obtainable by other methods. If we perform time sharing between all sessions and also between all the sources, the approximation factor to cut-set bound is of order SK. A smarter strategy is to choose one source per session to communicate to its corresponding sink, so that we get a multiple unicast problem, for which Leighton and Rao [90] showed a $O(\log K)$ gap between flows and cuts. Since we are time sharing between the S sources, we get a factor $O(S \log K)$ between our achievable strategy and the cut-set bound. In comparison to these results, we see that our approximation factor of $O(\log SK)$ is much stronger. Furthermore, we observe that a well provisioned network will have capacities scaling linearly with K, the number of sessions. Thus, the cut-set bound will scale as K and hence our achievable rates will scale at least on the order of $\frac{K}{\log SK}$.

Theorem 33 is proved by first showing a connection between Steiner packing rates and the cut-set bound (this connection has been established by prior work in the approximation algorithms literature [77, 108, 56]),

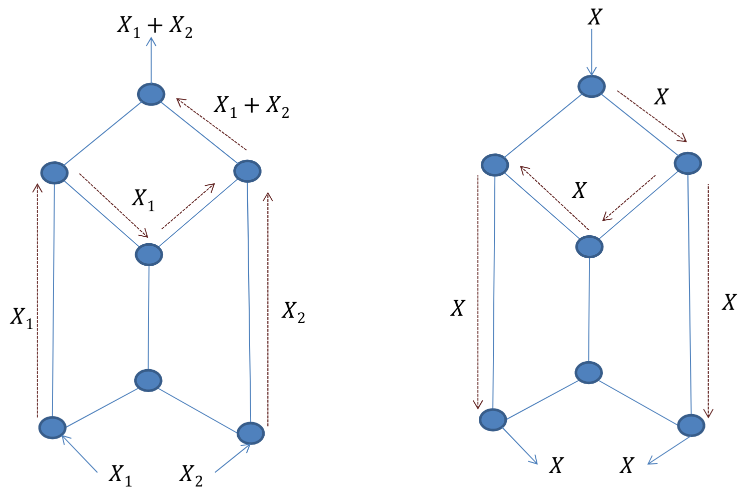
$$\frac{\bar{\mathcal{C}}}{g(S,K)} \subseteq \mathcal{R}_s \subseteq \bar{\mathcal{C}}, \tag{4.17}$$

where $\mathcal{R}_s$ is the Steiner packing rate region, and then using the fact that

$$\mathcal{R}_{\text{comp}} = \frac{\mathcal{R}_s}{\lambda_f(S)}. \tag{4.18}$$

4.4.1 Relation between computation and communication

Our results are strongly motivated by an analogy between the function computation problem and a multi-session communication problem, and analogous results available for the communication problem. This "dual" communication problem is obtained by reversing the nature of the sources and destinations. In particular, in a function computation session, a single sink node computes the function of many sources; in the communication problem, many destinations demand the same information from the single source (this is called multicasting). The key observation is that for the computation problem, every Steiner tree is a computation tree and for the communication problem, every Steiner tree is a multicasting



(a) A network for function computation

(b) A multicast network

Figure 4.3: Dual communication and computation networks

tree. In the computation problem, whenever information streams merge in the Steiner tree, in-network computation is performed. In the dual communication problem, whenever information streams bifurcate in the Steiner tree, replication is performed. Thus, viewed in this manner, *replication and computation are dual operations*. Therefore, the achievable rates have a natural relationship. It turns out that the cut-set bound for the computation problem in this thesis also has a natural analog for the communication problem.

Formally, given a function computation network, we will define the communication problem as the following: the communication graph G is the same with the same capacities. In the computation problem, there are K functions to be computed. In the communication problem, there are K independent messages to be communicated. For each message k, there is a source ρ^k and S destinations defined by the set S_k, each of which desire the message. The k messages are independent. This problem is called the *multiple multicasting problem* in the network coding literature. Thus, for every multi-session computation problem, there is a dual multiple multicast communication problem.

We refer the reader to Fig. 4.3 for an example of a function computation problem over an undirected butterfly network, and the dual problem of multicasting. In Fig. 4.3a, there are two sources with messages X_1 and X_2 and one destination demanding $X_1 + X_2$. A Steiner tree connecting the three nodes is also shown in the figure with dotted lines. Till the two streams meet, the messages X_1 and X_2 are passed along separately, and when they meet the function $X_1 + X_2$ is computed and forwarded along. The dual communication problem in Fig. 4.3b has a single source X which needs to be multicast to two destinations. Again, the Steiner tree is shown in dotted lines; in this multicast problem, when the Steiner tree bifurcates, the information X is replicated on both the outgoing edges.

Achievable strategy: We will now see that both our achievable strategy and the cut-set bound for the computation problem have a natural analog in the communication problem as well. First, we focus on the achievable strategy. For the computation problem, our achievable strategy is based on packing Steiner trees for the various G_k. For the communication problem also, packing Steiner trees forms a natural achievable strategy, since a Steiner tree can be used to disseminate a message from one of the nodes in G_k, namely ρ_k to the rest of the nodes S_k. Thus given any fractional Steiner packing $(\pi_1, ..., \pi_K)$, a rate tuple $(\pi_1, ..., \pi_K)$ can be achieved for the multiple-multicasting problem. Let $\mathcal{R}_{\text{m.m.}}$ denote the achievable region for the multiple multicast problem. By the above observation, this equals the Steiner packing region, i.e., $\mathcal{R}_{\text{m.m.}} = \mathcal{R}_s$.

Outer bound: Given any set $\Omega \subseteq V$, define $K(\Omega)$ as before $K(\Omega) := \{k : G_k \cap \Omega \neq \emptyset, G_k \cap \Omega^c \neq \emptyset\}$ as the set of sessions disconnected by Ω and define $\text{Cut}(\Omega) := \sum_{(ij)\in E: i\in\Omega, j\in\Omega^c} c(ij)$. This implies that if we separate Ω and Ω^c, for each $k \in K(\Omega)$, at least one destination is separated from the source and thus

$$\bar{\mathcal{C}} = \{\overline{R} : \sum_{k\in K(\Omega)} R_k \leq \text{Cut}(\Omega)\ \forall \Omega\} \tag{4.19}$$

is an outer bound on the set of rates achievable for the multiple-multicast problem.

Capacity approximation: Since the achievable strategy is given by Steiner packing and the outer bound is given by cut-set bound, we can now bound the gap between the two as

$$\frac{\bar{\mathcal{C}}}{g(S, K)} \leq \mathcal{R}_{\text{m.m.}} \leq \bar{\mathcal{C}}, \tag{4.20}$$

where $g(S, K)$ is as defined in Theorem 33. Thus we see that the function computation problem and the communication problem have a natural duality for the proposed achievable strategy and the outer bound.

Linear function computation: Consider the special case when the source alphabet $\mathcal{A}$ is a field and the function to be computed is linear over the field. The linear function has $\lambda_f = 1$ and is marginally injective (see Sec. 4.2.2). Thus we have

$$\frac{\bar{\mathcal{C}}}{g(S,K)} \subseteq \mathcal{R}_{\text{comp}} \subseteq \bar{\mathcal{C}}. \tag{4.21}$$

For the case of a single session $K = 1$, $g(S, K) \leq 2$, which implies that the achievable rate region using Steiner packings and the cut-set bound are approximately the same, to within a factor of 2. This is the dual of the corresponding result for multicasting [92], which shows that Steiner packing and cut are within a factor of 2, thus bounding the gain due to network coding in undirected graphs.

The single-session linear-function-computation problem has been well studied, and for directed graphs (potentially cyclic), it is known that the cut-set bound is achievable [127]. The achievable strategy is given by random linear coding and is inspired by the strategy used for the dual multicasting problem. The duality between linear coding for single-session linear-function computation and linear coding for multicasting was observed in [127]. However, there is no natural way to extend this duality (or even the achievable strategy) to the computation of general functions. Furthermore, our strategy of Steiner packings enables us to tackle the case of multi-session linear-function computation and show a provable approximation ratio of $O(\log SK)$ between the achievable strategy and the cut-set bound.

4.4.2 Proof of Theorem 33 for the special case, $K = 1$

In this setup, one node wants to compute a function of all the sources. Since $K = 1$, the regions collapse to single numbers, for which we use small case letters. There is a close relationship between the fractional Steiner packing number $R_s = \pi$ and the cut $\bar{C}$ (which is also called the Steiner cut). This relationship is based on the Tutte-Nash-Williams theorem [144, 110] and Mader's undirected splitting-off theorem [47] and was elucidated in the multicast setting by Li and Li [92]:

$$\frac{1}{2}\bar{C} \leq R_s \leq \bar{C}. \tag{4.22}$$

Therefore, by using this fractional Steiner packing, and $R_{\text{comp}} = \frac{1}{\lambda_f} R_s$, we get the desired result,

$$\frac{1}{2\lambda_f}\bar{C} \subseteq R_{\text{comp}}. \tag{4.23}$$

4.5 Proof of Main Result

We will now prove this result for general K using connections to algorithmic work showing approximation algorithms for "sparsest Steiner cuts." While the flow-cut gap for Steiner flows and cuts is already available in the approximation algorithms literature [77, 108, 56], we will proceed to give a tutorial overview of the key steps involved in this proof for the sake of readers with a different background.

4.5.1 Description of rate regions

For the case of multiple sessions, we have to deal with rate regions and cut-set regions. In order to deal with these regions in a compact manner, we use the max-concurrent flow representation. Let $(D_1, ..., D_K)$ be given demands of sessions 1,..,K. We want to achieve a rate proportional to the given demands, i.e.,

$$(R_1, ..., R_K) = \alpha(D_1, ..., D_K). \tag{4.24}$$

In this representation, we would like to find the maximum value of α such that $\alpha(D_1, ..., D_K)$ is in the capacity region $\mathcal{C}$. The achievable rate is given by $\frac{1}{\lambda_f}(\pi_1, ..., \pi_K)$, where $(\pi_1, ..., \pi_K)$ is a simultaneous fractional Steiner packing. We set $(\pi_1, ..., \pi_K) = \gamma(D_1, ..., D_K)$ and want to maximize γ, this problem is called the maximum concurrent Steiner flow problem. The maximum value γ^* is called the (maximum concurrent) Steiner packing rate.

The maximum concurrent Steiner flow problem can be written as

$$\gamma^* = \max \gamma \quad \text{s.t.} \tag{4.25}$$

$$\sum_{\tau \in \mathcal{T}_k} f_\tau \geq \gamma D_k \quad \forall k \in [K] \tag{4.26}$$

$$\sum_k \sum_{\tau \in \mathcal{T}_k : e \in \tau} f_\tau \leq c(e) \quad \forall e \in E. \tag{4.27}$$

The cut-set bound on $\mathcal{C}$ now translates to

$$\alpha \leq \frac{\text{Cut}(\Omega)}{D(\Omega)} =: \nu(\Omega), \tag{4.28}$$

where we refer to $\nu(\Omega)$ as the sparsity of the cut Ω. Thus

$$\alpha \leq \nu^* := \min_{\Omega \subseteq V} \nu(\Omega), \tag{4.29}$$

where the minimizer Ω^* is called the *sparsest Steiner cut*. The cut-set bound also bounds the Steiner packing rate and hence, $\gamma^* \leq \nu^*$ (see [77]).

We note that while we have defined cuts using a subset of the vertex set $\Omega \subseteq V$, there is an alternate way of defining the cuts using subsets of edges $F \subseteq E$. We define the sparsity of edge cut F as

$$\nu(F) = \frac{\text{Cut}(F)}{D(F)}, \tag{4.30}$$

where $\text{Cut}(F) = \sum_{f \in F} c(f)$ and $D(F) = \sum_{k \in K(F)} D_k$, and $K(F)$ is the set of sessions separated by F, i.e., at least one node of G_k is disconnected from another node of G_k in the graph $(V, E \setminus F)$. In general, cuts based on edge sets and vertex sets can be very different, but for undirected graphs, the two turn out to be equivalent (see [24]), i.e.,

$$\nu^* = \min_{\Omega \subseteq V} \frac{\text{Cut}(\Omega)}{D(\Omega)} = \min_{F \subseteq E} \frac{\text{Cut}(F)}{D(F)}. \tag{4.31}$$

4.5.2 Dual of the Steiner flow problem

We would like to show provable bounds on the ratio between γ^* (Steiner packing rate) and ν^* (the cut). This can then be translated into bounds on the ratio between α (computation rate) and ν^* (the cut). In order to do this, we first write the dual of the linear program for γ^*. There is a dual variable y_k for each k, corresponding to the constraints in (4.26) and a dual variable ℓ_e for each $e \in E$. We note that we can treat the graph as fully connected without loss of generality (since the capacity of non-existent edges can be set to zero). The dual can be written as follows:

$$\begin{aligned} \gamma^* &= \min \sum_{e \in E} c_e \ell_e \quad \text{s.t.} \\ \sum_{e: e \in \tau} \ell_e &\geq y_k \quad \forall \tau \in \mathcal{T}_k \, \forall k \in [K] \\ \sum_{k \in [K]} D_k y_k &\geq 1. \end{aligned}$$

Let $w_\ell(\tau)$ be the weight of the tree τ with weights ℓ, i.e.,

$$w_\ell(\tau) = \sum_{e: e \in \tau} \ell_e. \tag{4.32}$$

Furthermore we define $w_\ell(G_k)$ as the minimum weight of the Steiner tree with nodes G_k:

$$w_\ell(G_k) = \min_{\tau \in \mathcal{T}_k} w_\ell(\tau). \tag{4.33}$$

With this notation the y_k in the optimization problem is set equal to $w_\ell(G_k)$ and therefore can be rewritten as follows:

$$\begin{aligned} \gamma^* &= \min \sum_{e \in E} c_e \ell_e \quad \text{s.t.} \\ \sum_{k \in [K]} D_k w_\ell(G_k) &\geq 1. \end{aligned} \tag{4.34}$$

4.5.3 Tree networks

We will first consider the case when the network graph is an undirected tree. We would like to show that $g(S, K) = 1$. Since the graph is a tree, there is only one Steiner tree τ_k for each set G_k, i.e.,

$$\mathcal{T}_k = \{\tau_k\} \quad \forall k \in [K]. \tag{4.35}$$

Therefore,

$$w_\ell(G_k) = w_\ell(\tau_k) \quad \forall k \in [K]. \tag{4.36}$$

Thus the dual program for Steiner tree packing can be rewritten as

$$\begin{aligned} \gamma^* &= \min \sum_{e \in E} c_e \ell_e \quad \text{s.t.} \\ \sum_{k \in [K]} D_k w_\ell(\tau_k) &\geq 1. \end{aligned} \tag{4.37}$$

We would like to compare the optimal value of this program to the sparsest cut. We start with the optimal solution $\ell(e)\ \forall e \in E$ for this dual program and then try to obtain a cut Ω whose sparsity $\nu(\Omega)$ is close to the value of this dual program.

When the graph is a tree, we show that there is a sparsest cut among the cuts that remove a single edge. For an edge e, we define $D(e)$ as the total sum of demands separated by removing the edge e (similar to the definition of $D(\Omega)$). We write $e|G_k$ to denote that edge e disconnects at least one node of G_k from another node of G_k. Thus

$$D(e) = \sum_{k \in [K]} \mathbb{1}_{[e|G_k]} D_k. \tag{4.38}$$

In this notation we can also write the Steiner tree τ_k for G_k as

$$\tau_k = \{e : e|G_k\} \quad \forall k \in [K], \tag{4.39}$$

since τ_k includes every edge that separates G_k.

We start with the optimal dual variables $\ell(e), e \in E$, which is feasible for the dual program, and write

$$\begin{aligned} \nu^* &= \min_{F \subseteq E} \frac{\text{Cut}(F)}{D(F)} \leq \min_{e \in E} \frac{c(e)}{D(e)} = \min_{e \in E} \frac{\ell_e c(e)}{\ell_e D(e)} \\ &\overset{(a)}{\leq} \frac{\sum_{e \in E} \ell_e c(e)}{\sum_{e \in E} \ell_e D(e)} \\ &= \frac{\sum_{e \in E} \ell_e c(e)}{\sum_{e \in E} \ell_e \sum_{k \in [K]} \mathbb{1}_{[e|G_k]} D_k} \end{aligned}$$

$$
\begin{aligned}
&\stackrel{(b)}{=} \frac{\sum_{e\in E}\ell_e c(e)}{\sum_{k\in[K]} D_k\{\sum_{e\in\tau_k}\ell_e\}} \\
&= \frac{\sum_{e\in E}\ell_e c(e)}{\sum_{k\in[K]} D_k w_\ell(\tau_k)} \\
&\stackrel{(c)}{\leq} \sum_{e\in E}\ell_e c(e) \\
&\stackrel{(d)}{=} \gamma^*,
\end{aligned}
$$

where (a) follows due to the standard inequality

$$\min_i \frac{a_i}{b_i} \leq \frac{\sum_i a_i}{\sum_i b_i}, \tag{4.40}$$

(b) follows from (4.39), the inequality (c) follows because ℓ is feasible for the dual program and hence satisfies the constraints in (4.37),

$$\sum_{k\in[K]} D_k w_\ell(\tau_k) \geq 1, \tag{4.41}$$

and (d) is immediate from (4.37) as well.

Since $\gamma^* \leq \nu^*$ always, we have that $\gamma^* = \nu^*$ and the value of the sparsest cut is equal to the maximum concurrent Steiner flow, if the graph is a tree.

4.5.4 General network

Now, we move on to considering a general network. For a general network, the following result is known [77]:

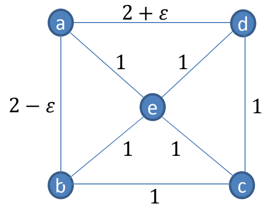
$$\frac{1}{O(\log^2 SK)}\nu^* \leq \gamma^* \leq \nu^*, \tag{4.42}$$

where the function $O(\log^2 SK)$ does not depend on the number of nodes, capacity of edges in the network or the demands $D_1, ..., D_K$. The approximation factor was refined to $O(\log n)$ in [108] (see Section 3.1 there) and then to $O(\log|\cup_k G_k|)$ in [56]. Note that $|\cup_k G_k| \leq SK$. Therefore, we can write

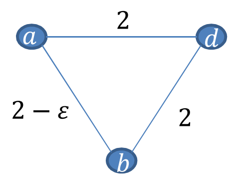
$$\frac{1}{O(\log SK)}\nu^* \leq \gamma^* \leq \nu^*. \tag{4.43}$$

This result is implicit in [108] and [56], and we refer the reader there for a detailed proof. We will give a sketch of the proof here for completeness. The basic idea of the proof is to connect Steiner flows and cuts in general networks to Steiner flows and cuts in tree networks. This is done using the notion of embeddings of general metric spaces into "tree metrics."

Algorithm for finding Steiner packings



(a) A weighted network



(b) Reduced graph

Figure 4.4: Restricted Steiner trees

First, we observe that the Steiner tree packing problem in general graphs (even for the case of a single session) is NP-hard [154]. We will describe a polynomial time algorithm, proposed in [77], for computing Steiner tree packing whose packing number is within a factor of two of the optimal Steiner tree packing. The *ellipsoid method* [154] gives a way of converting optimization problems into feasibility checking. In particular, given an assignment of the variables, if there is a polynomial time algorithm (called the *separation oracle*) that will either say it is feasible or produce a separating hyperplane that separates the feasible set and the assigned variables, then we can use this separation oracle to do optimization in polynomial time as well.

We will apply the ellipsoid method to the dual program of Steiner tree packing as formulated in (4.34). Given an assignment of $\ell(e)$, checking the feasibility is very easy once we compute $w_\ell(G_k)$, i.e., the minimum weight of the Steiner tree with nodes G_k. This minimum weight Steiner tree problem is NP-hard, but there is a factor 2 approximation algorithm that runs in polynomial time [154]. We briefly describe this algorithm. Given a set of distances $\ell(e)$, create a graph H_k, which has vertices G_k and is fully connected. The distance between two nodes u and v in H_k is the distance of the shortest path P_{uv} between u and v in the original graph G with distance $\ell(e)$ on edge $e = uv$. Now, we find a minimum weight spanning tree with edges E_k in H_k using Prim's or Kruskal's algorithm [30]. This yields a Steiner tree τ_k with edge set $\cup_{uv:(uv)\in E_k} P_{uv}$ on the original graph G (we may have to delete certain edges to get a tree). Steiner trees obtained in this manner are called restricted Steiner trees. It can be shown that the weight of the restricted Steiner tree, denoted by $\tilde{w}_\ell(G_k)$, is within a factor 2 of the minimum weight Steiner tree, i.e.,

$$w_\ell(G_k) \leq \tilde{w}_\ell(G_k) \leq 2w_\ell(G_k). \tag{4.44}$$

For an example of the minimum weight Steiner tree problem, see Fig. 4.4. A weighted graph is shown in Fig. 4.4a, where we would like to construct a minimum weight Steiner tree with nodes a,b, and d. To do so, we construct a graph H on nodes a,b and d with weights equal to the minimum distance between these nodes in the original graph. This graph is shown in Fig. 4.4b. The minimum weight spanning tree in H is given by the edges ba and ad of weight $4-\epsilon$. This corresponds to a tree in the original graph with the edges ba, ae and ed. The minimum weight Steiner tree in the original graph is of weight 3, given by the edges be, ae and ed.

Since the objective function is linear, using this approximation algorithm for minimum weight Steiner tree as the separation oracle in ellipsoid method yields a packing which is within a factor 2 of the best packing. We call this program the restricted Steiner flow problem:

$$\begin{aligned} \min \sum_{e\in E} c_e \ell_e \quad & \text{s.t.} \\ \sum_{k\in[K]} D_k \tilde{w}_\ell(G_k) \quad & \geq \quad 1. \end{aligned} \tag{4.45}$$

We will first show that there is an optimal solution to the restricted Steiner flow problem such that $\ell(uv)$ is a semi-metric, i.e., it satisfies the triangle inequality,

$$\ell(uv) \leq \ell(uw) + \ell(wv) \quad \forall u, v, w \in V. \tag{4.46}$$

Given a graph G and distances $d(e)$ for each edge, $e \in E$, we can define the shortest path metric $d_G(uv)$ as the shortest distance between u and v, i.e.,

$$d_G(uv) := \min_{p\in\mathcal{P}_{uv}} \sum_{e:e\in p} d(e) \quad \forall u, v \in V, \tag{4.47}$$

where $\mathcal{P}_{uv}$ is the set of paths between nodes u and v. Given an optimal dual solution $\ell(e)$ for the graph G, we define $\hat{\ell} = \ell_G$. Now $\hat{\ell}(e)$ also satisfies the constraints since $\tilde{w}_\ell$ depends only on pair-wise distances, which remains the same for ℓ and $\hat{\ell}$. Furthermore, the objective function does not increase because $\hat{\ell}(e) \leq \ell(e)\ \forall e \in E$ and reducing the ℓ only reduces the objective function. Thus there exists an optimal solution which is a semi-metric.

Tree embeddings

In order to "convert" the general network problem into a tree network problem, we use the notion of tree embeddings. We start with the definition of a tree metric (we refer the reader to [24] for a detailed discussion).

Definition 3. *Given a tree $T = (V, E_T)$ and distances $d(e), e \in T$, the induced shortest path metric $d_T(uv)$ on the set of nodes V, is called a tree metric.*

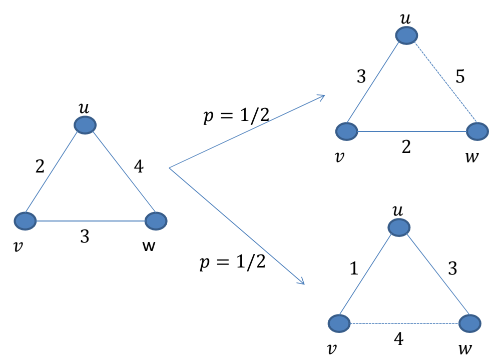


Figure 4.5: Tree embedding

Thus, given weights $d(e)$ on the edges of a tree T, we can obtain a tree metric d_T on the nodes.

Lemma 30. *[42], [56] Given a complete undirected graph $G = (V, E)$ with a metric ℓ on V, there is a randomized polynomial time algorithm that produces a random edge weighted tree $T = (V, E_T)$ inducing a tree metric ℓ_T such that*

1. $\ell(uv) \leq \ell_T(uv)\ \forall u, v \in V$,
2. $\mathbb{E}(\ell_T(uv)) \leq O(\log SK)\ell(uv)\ \forall u, v \in V$.

Proof. This result was proved with a factor $O(\log n)$ in [42] and this factor was improved to $O(\log SK)$ in [56]. $\square$

For an example of a tree embedding, see Fig. 4.5, where a graph G with nodes u, v, w is embedded into tree T_1, comprised of edges uv and vw (the dotted edge uw is not present) with probability $\frac{1}{2}$ and into T_2 comprised of edges vu and vw with probability $\frac{1}{2}$. This embedding is exact, because the average distance for each pair of nodes turns out to exactly equal the original distances in the graph G.

Steiner flow-cut gaps

We start with the restricted Steiner flow problem, see (4.45), with optimal solution $\ell(e)$. We would like to compare the optimal value of this problem to the sparsest Steiner cut problem. We know that in a tree, Steiner flow and sparsest cut are close. Therefore, we use Lemma 30 for embedding the optimal solution ℓ of the restricted Steiner flow problem into a tree randomly. Let T, ℓ_T denote the random tree and the tree metric into which the given metric ℓ is embedded. In a tree T, every cut comprises of a single edge $e \in E_T$. Corresponding to this cut $e \in E_T$, there is a cut in the original graph defined by

$$\text{Sep}_e := \{(i, j) | i, j \in V, e \text{ separates } i, j \text{ in } T\}. \tag{4.48}$$

Removing the edges Sep_e in G disconnects at least one node of G_k from another node of G_k in G. The value of the cut, denoted by $c(\text{Sep}_e)$ is defined as

$$c(\text{Sep}_e) = \sum_{\tilde{e} \in \text{Sep}_e} c(\tilde{e}). \tag{4.49}$$

As before, we define $e|G_k$ to denote that removing edge e disconnects at least one node of G_k from another node of G_k in T.

We start with optimal dual variables $\ell(e), e \in E$ for the restricted Steiner flow program. We can write a set of inequalities similar to the tree networks case.

$$
\begin{aligned}
\nu^* &= \min_{F \subseteq E} \frac{\text{Cut}(F)}{D(F)} \leq \min_{e \in E_T} \frac{c(\text{Sep}_e)}{D(\text{Sep}_e)} \\
&= \min_{e \in E_T} \frac{\ell_T(e) c(\text{Sep}_e)}{\ell_T(e) D(\text{Sep}_e)} \\
&\overset{(a)}{\leq} \frac{\sum_{e \in E_T} \ell_T(e) c(\text{Sep}_e)}{\sum_{e \in E_T} \ell_T(e) D(\text{Sep}_e)} \\
&= \frac{\sum_{e \in E_T} \ell_T(e) \sum_{(u,v) \in \text{Sep}_e} c(uv)}{\sum_{e \in E_T} \ell_T(e) \sum_{k=1}^{K} \mathbb{1}_{[e|G_k]} D_k} \\
&= \frac{\sum_{u,v} c(uv) \sum_{e \in E_T : (u,v) \in \text{Sep}_e} \ell_T(e)}{\sum_k D_k \sum_{e \in E_T : e|G_k} \ell_T(e)} \\
&= \frac{\sum_{u,v} c(uv) \ell_T(uv)}{\sum_k D_k \sum_{e \in E_T : e|G_k} \ell_T(e)} \\
&\leq \frac{\sum_{u,v} c(uv) \ell_T(uv)}{\sum_k D_k \sum_{e \in E_T : e|G_k} \ell(e)} \\
&\overset{(b)}{\leq} \frac{\sum_{u,v} c(uv) \ell_T(uv)}{\sum_k D_k w_\ell(T)} = \frac{\sum_{u,v} c(uv) \ell_T(uv)}{\sum_k D_k w_\ell(G_k)} \\
&\overset{(c)}{\leq} 2 \frac{\sum_{u,v} c(uv) \ell_T(uv)}{\sum_k D_k \tilde{w}_\ell(G_k)} \\
&\overset{(d)}{\leq} 2 \sum_{u,v} c(uv) \ell_T(uv), \\
\Rightarrow \nu^* &\leq 2\mathbb{E}\left[\sum_{u,v} c(uv) \ell_T(uv)\right] \\
&= 2\left(\sum_{u,v} c(uv) \mathbb{E}[\ell_T(uv)]\right) \\
&\overset{(e)}{\leq} 2O(\log SK)\left(\sum_{u,v} c(uv) \ell(uv)\right) \\
&\overset{(f)}{=} O(\log SK)\gamma^*,
\end{aligned}
$$

where (a) follows due to the standard inequality (4.40). (b) follows from the fact that the set of edges $\{e \in E_T : e|G_k\}$ is a Steiner tree on G_k and therefore its weight is greater than the minimum weight spanning tree on G_k. (c) follows from the fact that

$$w_\ell(G_k) \leq 2\tilde{w}_\ell(G_k) \quad \forall k \in [K]. \tag{4.50}$$

(d) is a direct consequence of the fact that ℓ is a feasible solution for restricted Steiner flow and hence satisfies the constraint given in (4.34),

$$\sum_{k\in[K]} D_k \tilde{w}_\ell(G_k) \geq 1. \tag{4.51}$$

The inequality (e) follows from Lemma 30 and (f) follows from the definition of γ^* in (4.34).

This proves the relationship between the Steiner packing and cuts

$$\frac{\nu^*}{O(\log SK)} \leq \gamma^* \leq \nu^*. \tag{4.52}$$

Approximation for computation capacity: Now, we can use the relationship between Steiner packing and cuts in order to derive approximations for function computation capacity. Using (4.18), we get $\alpha = \frac{\gamma^*}{\lambda_f}$. This implies that, for computation of λ_f-divisible functions,

$$\frac{1}{\lambda_f O(\log SK)} \nu^* \leq \alpha. \tag{4.53}$$

Since this approximation factor $O(\log SK)$ between the achievable rate and the cut does not depend on the demand vector $D_1, ..., D_K$, we can show that this approximation factor holds for the entire rate region,

$$\frac{1}{\lambda_f O(\log SK)} \bar{\mathcal{C}} \subseteq \mathcal{R}_{\text{comp}}, \tag{4.54}$$

which proves the achievable portion of the result.

CHAPTER 5

CONCLUSIONS AND FUTURE WORK

> "We are trying to prove ourselves wrong as quickly as possible, because only in that way can we find progress." - Richard Feynman

In this thesis, we proposed a layered architecture for wireless networks and developed new techniques to analyze the performance of such an architecture. We showed that powerful techniques developed in theoretical computer science for understanding communication over wireline network graphs can be generalized to polymatroidal networks. We also showed that polymatroidal networks form a network-layer interface for a wireless network, i.e., any scheme in the polymatroidal network can be translated to the wireless network. Furthermore, we demonstrated that similar results can be obtained for the problem of computation of a function of independent sources in a capacitated graph at multiple terminals.

We believe that this research opens up several interesting research directions, some of which we survey below.

5.1 Wireless Networks

A main result of this thesis is that good physical schemes, which achieve close to the cut-set bound and obey a certain reciprocity, can be used in a natural layered architecture to obtain schemes for arbitrary networks composed of the corresponding channels. Furthermore, this layered architecture achieves the capacity to within a logarithmic factor, thus showing that the design of an optimal network communication scheme boils down to the design of good physical layer schemes. In particular, in the case of a general wireless network, interference alignment schemes for the (physical-layer) interference channel formed the basis. However, existing interference alignment schemes suffer from several drawbacks when it comes to practical implementation. Thus one question that needs to be addressed in order to make this work practical is the following:

Open Question 1**:** *Can we design better physical layer schemes for the interference channel?*

Following the hints offered by the erasure channel model, the inclusion of the natural feedback present in the interference channel may create the possibility of obtaining practical interference alignment schemes for this problem. We have obtained wireless network capacity results to within a logarithmic factor in the number of source-destination pairs. In fact, unless network coding or better-than-cut bounds are used, this factor cannot be bettered even for the special case of wireline networks. A conjecture in wireline network theory called the Li-Li conjecture [92], states that routing is optimal for undirected wireline networks. Our work provides a new impetus to solve the Li-Li conjecture, since the resolution of this conjecture may have ramifications for a much wider class of networks.

Open Question 2**:** *What is the ramification of the resolution of Li-Li conjecture for wireless networks?*

The capacity results in this thesis have been primarily for the case when the wireless channels are symmetric, the so called bi-directed wireless network model. However, for the polymatroidal network model, we saw that even if the communication network is directed, as long as the traffic model is symmetric, i.e., if s_i is communicating information to t_i at rate R_i, then t_i is also needed to communicate information to s_i at rate R_i, we showed poly-logarithmic flow-cut gaps. It turns out that the edge-cuts involved in this problem do not correspond to the information-theoretic cut-set bound. Using more general outer bounds, it was shown recently that these bounds are indeed fundamental, i.e., they bound the rate of any communication scheme for wireline networks. In collaboration with others [67], we have shown that these results in fact carry over to wireless networks as well.

Without this symmetry assumption for either the traffic or for the channel, it is not clear whether we can obtain interesting results for wireless networks. However, this setting for even wireline networks is wide open. In this setting, one possible scheme involves i.i.d. random coding at the intermediate nodes which induces an end-to-end fast fading interference channel between the transmitters and the receivers, for which one can apply interference alignment techniques.

Open Question 3**:** *How good is random coding along with end-to-end interference alignment for directed wireline networks?*

While we focused on worst case bounds in this thesis, some of the encountered wireless networks may have additional structure. One way of imposing structure is to consider random graphs, for example, Erdos-Renyi random graphs, where any two nodes are connected with a probability p and each node has a message to every other node, with the number of nodes tending to infinity. In the case of a wireline network, there is a constant flow-cut gap and nice structural results on optimal flows under this model [6]. So a natural question is whether there are corresponding counterparts for wireless networks.

Open Question 4**:** *Can we obtain constant-factor capacity approximations for* random *wireless networks?*

The communication schemes proposed for small canonical channels and the communication schemes proposed for large networks in the scaling-law literature seem to be disparate. It is known that when the network is dense, then hierarchical MIMO [121] is optimal whereas when the network is extended over a large scale, multi-hopping [55] is optimal. There are also various regimes in the middle where a hybrid scheme involving hierarchical MIMO at local scales and multi-hopping over larger scales is needed [120]. This architecture is reminiscent of our local physical-layer plus global-routing scheme. For the case of a dense network where only local physical layer schemes are necessary, recent work [114] has made this connection by using the ergodic interference alignment scheme [112] as the building block. Thus, the following question begs to be asked:

Open Question 5**:** *Can the layered architecture directly yield scaling law results for geographical wireless networks?*

The results of this thesis were obtained under a general traffic pattern of multiple-unicasting. However, as the use of wireless data increases, many users are watching the same videos simultaneously, particularly during a high coverage event like a sports game or elections. In this context, it has been observed that sending the same information through distinct unicast streams is inherently not scalable. Hence, it is important to design simple schemes for more general traffic models, the most general being multiple-multicasting, where each information is requested by a subset of users. This leads us to the following question:

Open Question 6**:** *Does the approximate optimality of layered architectures generalize to more complex traffic patterns like multiple-multicasting?*

5.2 Polymatroidal Networks

As we have shown, polymatroidal networks act as a network-layer interface for wireless networks. Thus, polymatroidal networks provide a new platform in which to study various scheduling and routing problems.

Open Question 7**:** *How far do standard techniques for routing and scheduling extend to polymatroidal networks?*

For planar wireline networks, Gupta et al. [57] conjectured that the concurrent flow-sparsest cut gap is $O(1)$ for the edge-capacitated setting. Rao [129] proved an upper bound of $O(\sqrt{\log n})$, thereby improving upon the gap for general graphs which can be $\Omega(\log n)$ in the worst case. The throughput flow-multicut gap is however known to be $O(1)$ [140] in such graphs; in this thesis, we showed that this extends to planar polymatroidal networks as well. This leads us to the following question.

Open Question 8**:** *Is there a constant concurrent flow-sparsest cut gap for planar polymatroidal networks?*

Linking systems [132] can be thought of as bi-polymatroidal networks, which is a generalization of polymatroidal networks. Linking systems have turned out to be useful to study unicast information flow in wireless networks [17, 8, 160, 52, 124, 71].

Open Question 9**:** *Can the results for multiple-unicast in polymatroidal networks be generalized to the linking systems model? Would these provide additional insights for the resolution of capacity of a broader class of wireless networks?*

5.3 Function Computation

In this thesis, we studied computation of multiple functions of independent data in undirected graphs. We proposed a simple strategy for computation, based on packing Steiner trees and performing in-network computation along the Steiner tree. We showed that for a wide class of functions, the achievable strategy and the proposed outer bound are close by showing an approximation factor which is the product of λ_f (which is a property of the function class f) and a factor logarithmic in the number of nodes involved in the computation. For function classes which have a large λ_f, this strategy is clearly not optimal. A large λ_f implies that different links in the computation tree use widely varying amounts of capacity. However, in our analysis and algorithms, we have only dealt with unit capacity computation trees (Steiner trees), and therefore our algorithms and analysis take a huge performance hit. One possible research direction is to model the varying amounts of information conveyed by the different links in a computation tree. We can then design new algorithms for packing Steiner trees, which have differing capacity constraints on distinct links. In particular, we can use certain sub-modularity properties and the structure of the function in order to both design these algorithms and to obtain tighter gaps between these polymatroidal Steiner packings and generalized cut bounds.

Open Question 10**:** *How do we pack Steiner trees with differing capacities on different edges, where these capacities come from some sub-modular function? How close are these Steiner tree packings to the corresponding cuts?*

APPENDIX A

PROOFS FOR CHAPTER 2

A.1 The Dual of Flow Problem

Lemma 31. *For a polymatroidal network, the dual of the maximum throughput flow problem is equivalent (in terms of value) to the program given in Fig. 2.1.*

Proof. We will show the proof for the undirected case, the proof for the directed case is similar. The program for maximum throughput flow is given by:

$$\max \quad \sum_{i} \sum_{p \in \mathcal{P}_{(s_k,t_k)}} f(p)$$

s.t.

$$\begin{aligned} \sum_{e:e\in S} \sum_{p:e\in p} f(p) &\leq \rho_v(S) \quad \forall S \subseteq \delta(v)\ \forall v \in V \\ f(p) &\geq 0 \quad \forall p \in \mathcal{P}_{(s_i,t_i)}, \forall i = 1 \dots k. \end{aligned}$$

The dual of the flow linear program can now be written. Let the dual variables $d_v(S_v)$ correspond to the non-trivial constraint in the above linear program. Then the dual linear program is:

$$\mathcal{P}_d := \min \quad \sum_{v \in V} \sum_{S \subseteq \delta(v)} d_v(S)\rho_v(S)$$

s.t.

$$\begin{aligned} \sum_{e=uv:e\in p} \left(\sum_{S\subseteq\delta(u):e\in S} d_u(S) + \sum_{S\subseteq\delta(v):e\in S} d_v(S) \right) &\geq 1 \quad \forall p \in \mathcal{P}_{(s_k,t_k)} \text{ where } \quad e = uv \\ d_u(S) &\geq 0\ \forall u \in V\ \forall S \subseteq \delta(u). \end{aligned}$$

This can be rewritten equivalently as

$$
\begin{aligned}
\mathcal{P}_d \quad &:= \quad \min \quad \sum_{v \in V} \sum_{S \subseteq \delta(v)} d_v(S) \rho_v(S) \\
& \text{s.t.} \\
\ell(e) \quad &:= \quad \left(\sum_{S \subseteq \delta(u): e \in S} d_u(S) + \sum_{S \subseteq \delta(v): e \in S} d_v(S) \right) \\
\text{dist}_\ell(s_i, t_i) \quad &\geq \quad 1 \quad 1 \leq i \leq k \\
d_u(S) \quad &\geq \quad 0 \, \forall u \in V \, \forall S \subseteq \delta(u).
\end{aligned}
$$

Let us define new variables $\ell(e,u)$, $\ell(e,v)$ for each edge $e = uv$, and rewrite the linear program:

$$
\min \quad \sum_{v \in V} \sum_{S \subseteq \delta(v)} d_v(S) \rho_v(S)
$$

$$
\begin{aligned}
& \text{s.t.} \\
\ell(e) \quad &:= \quad \ell(e,u) + \ell(e,v), \text{ where } e = uv \\
\ell(e,u) \quad &= \quad \sum_{S \subseteq \delta(u): e \in S} d_u(S) \, \forall e \in E, e = uv \\
\ell(e,v) \quad &= \quad \sum_{S \subseteq \delta(v): e \in S} d_v(S) \quad \forall e \in E, e = uv \\
\text{dist}_\ell(s_i, t_i) \quad &\geq \quad 1 \quad 1 \leq i \leq k \\
d_u(S) \quad &\geq \quad 0 \\
\ell(e,u), \ell(e,v) \quad &\geq \quad 0 \quad \forall u \in V \, \forall S \subseteq \delta(u).
\end{aligned}
$$

The minimization is over the variables $\ell(e,u)$ and $d_v(S)$. Observe that for any fixed v the variables $d_v(S), S \subseteq \delta(v)$ influence only the variable $\ell(e,v), e \in \delta(v)$. Hence, for any v and a fixed assignment set of values $\ell(e,v), e \in \delta(v)$, the optimal choice of variables $d_v(S), S \subseteq \delta(v)$ can be obtained by solving the following linear program:

$$
\begin{aligned}
& \min \qquad \sum_{S \subseteq \delta(v)} d_v(S) \rho_v(S) \\
& \text{s.t.} \\
\sum_{S \subseteq \delta(v): e \in S} d_v(S) \quad &= \quad \ell(e,v) \quad \forall e \in E, e = uv \\
d_u(S) \quad &\geq \quad 0, \quad S \subseteq \delta(v), \, \forall v \in V.
\end{aligned}
$$

Recalling the definition of the convex closure of a function, one sees that the value of the above linear program is equal to $\tilde{\rho}_v(\mathbf{d}_v)$; note that for polymatroids we can drop the constraint $\sum_S d_v(S) = 1$ in the linear program for the convex closure. Since the convex closure is equal to the Lovász extension we obtain the desired equivalence of the formulations. $\square$

A.2 Proof of Lemma 8

Lemma 32. *Let $g : V \to [0, \beta]$ be a contraction, let $0 \le a_0 \le a < b \le b_0 \le \beta$ and $S_\theta = \{u \mid g(u) < \theta\}$. Suppose for every edge $e = uv \in \cup_{\theta \in [a,b]} \delta(S_\theta)$, $g(u)$ and $g(v)$ are both in $[a_0, b_0]$. Then,*

$$\int_a^b \nu(\delta(S_\theta)) d\theta \le 2 \sum_{v : g(v) \in [a_0, b_0]} \hat{\rho}_v(\mathbf{d}_v).$$

Proof. Consider an edge $uv \in \delta(S_\theta)$ and for simplicity assume $g(u) < g(v)$. The length of e in the embedding is $\ell'(e) = |g(v) - g(u)| \le \ell(e)$. The edge $(u, v) \in \delta(S_\theta)$ iff θ is in the interval $[g(u), g(v)]$. Also by the conditions of the theory for every such (u, v), $g(u) \in [a_0, b_0]$ and $g(v) \in [a_0, b_0]$. Note that the cost $\nu(\delta(S_\theta))$ is in general a complicated function to evaluate. We upper-bound $\nu(\delta(S_\theta))$ by giving an explicit way to assign $e = uv$ to either u or v as follows. Recall that in the relaxation $\ell(e) = \ell(e, u) + \ell(e, v)$ where $\ell(e, u)$ and $\ell(e, v)$ are the contributions of u and v to e. Let $r = \frac{\ell(e,u)}{\ell(e)}$ and let $\ell'(e, u) = r\ell'(e)$ and $\ell'(e, v) = (1 - r)\ell'(e)$. We partition the interval $[g(u), g(v)]$ into $[g(u), g(u) + \ell'(e, u))$ and $[g(u) + \ell'(e, u), g(v)]$; if θ lies in the former interval we assign e to u, otherwise we assign e to v. This assignment procedure describes a way to upper-bound $\nu(\delta(S_\theta))$ for each θ. Now we consider the quantity $\int_a^b \nu(\delta(S_\theta)) d\theta$ and upper bound it as follows.

Consider a node u and let $L_u = \{uv \in \delta(u) \mid g(v) < g(u)\}$ be the set of edges uv that go from u to the left of u in the embedding g. Similarly $R_u = \{uv \in \delta(u) \mid g(v) \ge g(u)\}$. Note that L_u and R_u partition $\delta(u)$. Let $\mathbf{d}'_u$ be the vector of dimension $|\delta(u)|$ consisting of the values $\ell'(e, u)$ for $e \in \delta(u)$. We obtain $\mathbf{d}^L_u$ from $\mathbf{d}'_u$ by setting the values for $e \in R_u$ to 0 and similarly $\mathbf{d}^R_u$ from $\mathbf{d}'_u$ by setting the values for $e \in L_u$ to 0. Since $0 \le \ell'(e, u) \le \ell(e, u)$ for each $e \in \delta(u)$ we see that $\mathbf{d}'_u \le \mathbf{d}_u$ and (component wise) and hence $\mathbf{d}^L_u \le \mathbf{d}_u$ and $\mathbf{d}^R_u \le \mathbf{d}_u$.

Since ρ_u is monotone we have that $\hat{\rho}_u(\mathbf{d}_u^L) \leq \hat{\rho}_u(\mathbf{d}_u)$ and $\hat{\rho}_u(\mathbf{d}_u^R) \leq \hat{\rho}_u(\mathbf{d}_u)$ (see Proposition 1).

We claim that

$$\int_a^b \nu(\delta(S_\theta))d\theta \leq \sum_{u \in V: g(u) \in [a_0, b_0]} (\hat{\rho}_u(\mathbf{d}_u^L) + \hat{\rho}_u(\mathbf{d}_u^R)),$$

which would prove the lemma.

In order to prove the claim, consider some fixed θ and $\nu(\delta(S_\theta))$. Fix a node u and consider the edges in $\delta(u) \cap S_\theta$ assigned to u by the procedure we described above; call this set $A_{\theta,u}$. First assume that $\theta < g(u)$. Then the edges assigned to u by the procedure, denoted by $A_{\theta,u} = \{e \in L_u \mid \theta > g(u) - \ell'(e,u)\}$. Similarly, if $\theta > g(u)$, $A_{\theta,u} = \{e \in L_u \mid \theta < g(u) + \ell'(e,u)\}$. From these definitions we have

$$\begin{aligned}
\nu(\delta(S_\theta)) &\leq \sum_{u \in V: g(u) \in [a_0, b_0]} \rho_u(A_{\theta,u}) \\
\Rightarrow \int_a^b \nu(\delta(S_\theta))d\theta &\leq \sum_{u \in V: g(u) \in [a_0, b_0]} \int_a^b \rho_u(A_{\theta,u})d\theta \\
&\leq \sum_{u \in V: g(u) \in [a_0, b_0]} \int_0^\beta \rho_u(A_{\theta,u})d\theta.
\end{aligned}$$

For a fixed node u,

$$\int_0^\beta \rho_u(A_{\theta,u})d\theta = \int_0^{g(u)} \rho_u(A_{\theta,u})d\theta + \int_{g(u)}^\beta \rho_u(A_{\theta,u})d\theta$$

Let $L_u = \{e_1, e_2, \ldots, e_h\}$ where $0 \leq \ell'(e_1, u) \leq \ell'(e_2, u) \leq \ldots \leq \ell'(e_h, u)$. Then

$$\int_a^{g(u)} \rho_u(A_{\theta,u})d\theta = \sum_{j=1}^h (\ell'(e_j, u) - \ell'(e_{j-1}, u))\rho(\{e_1, e_2, \ldots, e_j\})$$

The right-hand side of the above is by construction and the definition of the Lovász extension, equal to $\hat{\rho}_u(\mathbf{d}_u^L)$. Similarly, $\int_{g(u)}^\beta \rho_u(A_{\theta,u})d\theta = \hat{\rho}_u(\mathbf{d}_u^R)$. $\square$

A.3 Proof of Theorem 7

Theorem 7. *Given any edge cut for an undirected polymatroidal network, there exists a bi-partition cut whose sparsity is at most* 2 *times the sparsity of the edge cut. Furthermore, this factor is tight.*

Proof. Start with an edge cut F. This edge cut partitions the nodes into connected components (after the edges in the cut have been removed). This induces a natural "vertex multi-partition" $V = V_1 \uplus V_2 ... \uplus V_M$, and we can define the edge cut corresponding to a vertex multi-partition as

$$F^m(V_1, ..., V_M) := \{e = (u, v) : u \in V_l, v \in V_m, l \neq m\}, \tag{A.1}$$

where the superscript m stands for multi-partition. It is easy to see that $D(F^m) = D(F)$ and $\nu(F^m) \leq \nu(F)$.

Construct an undirected graph H with nodes $\hat{v}_1, ..., \hat{v}_M$ and edges $\hat{v}_i\hat{v}_j$ with weight w_{ij} equal to the demand between partition V_i and V_j in the original graph G. For graph H, there exists a *weighted max-cut*, whose value is greater than half the sum of all the weights (since a random bi-partition of H where each edge gets cut with probability half has expected weight equal to half the sum of all weights). Let this max-cut partition H into sets A and A^c. If we take the set $S = \cup_{i:\hat{v}_i \in A} V_i$ and S^c as a partition in the original graph G, the bi-partition cut F_S separates at least half the demand as the multi-partition. This implies that $D(F_S) \geq \frac{1}{2}D(F_m) \leq \frac{1}{2}D(F)$ and also $\nu(F_S) \leq \nu(F_m) \leq \nu(F)$, and therefore

$$\frac{\nu(F_S)}{D(F_S)} \leq 2\frac{\nu(F_S)}{D(F_S)}, \tag{A.2}$$

and we have obtained a bi-partition with at most twice the sparsity of the edge cut.

To see that this factor is tight, consider a polymatroidal network where there are n nodes v_0, v_1,...,v_{n-1}, with edge e_i between v_0 and v_i, for each $i \in \{1, 2, ..., n-1\}$, and let n be even. The only capacity constraint is a polymatroidal constraint at node v_0, which constrains the sum of every subset of $\{e_1, ..., e_{n-1}\}$ by a value of 1. The demand graph is a complete graph with each demand of unit value.

Now consider an edge cut F which removes all the edges. For such an F, $\nu(F) = 1$ and $D(F) = \binom{n}{2}$, so the sparsity is $\frac{2}{n(n-1)}$. On the other hand, any bi-partition cut F_S has $\nu(F_S) = 1$ and $D(F_S) = |S||S|^c$. The sparsest cut is one which maximizes $|S||S|^c$. This happens when $|S| = \frac{n}{2}$ and the sparsity of this cut is given by $\frac{4}{n^2}$. Thus the sparsest bi-partition cut is a factor of $\frac{2(n-1)}{n}$ bigger than the sparsity of the best edge-cut. This factor approaches 2 as n approaches ∞. $\square$

APPENDIX B

PROOFS FOR CHAPTER 3

B.1 Proof of Lemma 19

The rate region for cut-set under product distribution is given by:

$$\mathcal{R}^{\text{MAC}}_{\text{cut,product}}(P) = \left\{ R : \sum_{i \in S} R_i \le \log(1 + \sum_{i \in S} |h_i|^2 P) \right\}. \tag{B.1}$$

The rate region for cut-set under general distribution is given by:

$$\mathcal{R}^{\text{MAC}}_{\text{cut,general}}(P) = \left\{ R : \sum_{i \in S} R_i \le \log(1 + (\sum_{i \in S} |h_i|)^2 P) \right\}. \tag{B.2}$$

By the Cauchy-Schwarz inequality, we get

$$\left(\sum_{i \in S} |h_i| \right)^2 P \quad \le \quad \left(\sum_{i \in S} |h_i|^2 \right) dP, \tag{B.3}$$

which in turn implies

$$\mathcal{R}^{\text{MAC}}_{\text{cut,general}}(P) \subseteq \mathcal{R}^{\text{MAC}}_{\text{cut,product}}(dP). \tag{B.4}$$

We can similarly show that

$$\mathcal{R}^{\text{BC}}_{\text{cut,general}}(P) = \mathcal{R}^{\text{MAC}}_{\text{cut,product}}(dP). \tag{B.5}$$

Along with the equality in (3.17), this implies that

$$\mathcal{R}^{\text{MAC}}_{\text{cut,general}}(P) \subseteq \mathcal{R}^{\text{MAC}}_{\text{ach}}(dP). \tag{B.6}$$

We similarly get, using (3.21) and (B.5),

$$\mathcal{R}^{\text{BC}}_{\text{cut,general}}(P) \subseteq \mathcal{R}^{\text{BC}}_{\text{ach}}(dP). \tag{B.7}$$

B.2 Proof of Lemma 20

Without feedback the capacity of this erasure broadcast channel can be easily found. This is because this erasure broadcast channel is stochastically degraded [31], and the capacity is given by

$$\left\{(R_1, ..., R_d)| \sum_{i=1,2,..,d} R_i \leq 1-\epsilon\right\}. \tag{B.8}$$

The rate region can be achieved by time sharing between the individual links. We can compare this rate to the cut-set bound which is given by

$$R_{\text{cut}} = \left\{(R_1, ..., R_D)| \sum_{i\in J} R_i \leq 1-\epsilon^{|J|} \quad \forall J \subseteq \{1,2,..,d\}\right\}. \tag{B.9}$$

The ratio between the sum rate of the scheme and the cut-set bound is the factor

$$\frac{1-\epsilon}{1-\epsilon^d}, \text{ which} \quad \rightarrow \quad \frac{1}{d}, \text{ as } \epsilon \rightarrow 1. \tag{B.10}$$

As expected, the time-sharing region does not compare very favorably to the cut-set bound.

B.3 Proof of Lemma 22

Consider the rate region with feedback $\mathcal{R}_{\text{ach,fb}}$. We would like to know for what value of A does $A\mathcal{R}_{\text{ach,fb}} \supseteq \mathcal{R}_{\text{cut}}$. Let us take a point in $\mathcal{R}_{\text{cut}}$, we would like to know, for what value of A does this imply $\sum_{i=1,2,..,d} \frac{R_i}{1-\epsilon^i} \leq A$. This is equivalent to

$$A = \max \sum_{i=1}^{n} \frac{R_i}{1-\epsilon^i}, \tag{B.11}$$

such that

$$\sum_{i\in J} R_i \leq 1-\epsilon^{|J|} \quad \forall J \subseteq \{1,2,..,d\}. \tag{B.12}$$

This is a linear optimization over a polymatroid and the optimal solution is given by the greedy algorithm [36],

$$(R_1,...,R_d) = (1-\epsilon, \epsilon-\epsilon^2, ..., \epsilon^{n-1}-\epsilon^n), \tag{B.13}$$

and the optimal value of the objective function is

$$\sum_{i=1}^{n} \frac{\epsilon^{i-1}-\epsilon^i}{1-\epsilon^i}. \tag{B.14}$$

Lets examine the i-th term in this sum, substituting $x = \epsilon^{-1}$,

$$\frac{x-1}{x^i-1} = \frac{1}{1+x^1+..+x^{i-1}} \leq \frac{1}{i}. \tag{B.15}$$

Therefore the sum is upper bounded by

$$A \leq \sum_i \frac{1}{i} \leq \log d. \tag{B.16}$$

B.4 Feedback: Multiple Access Erasure Channel

Consider a finite field multiple access erasure channel, where

$$y = \sum_{i=1}^{d} e_i x_i, \tag{B.17}$$

where e_i are i.i.d. Bernoulli with probability $1-\epsilon$. In this channel, some of the transmitters' packets can get erased, and the received vector is the sum of those packets that did not get erased.

This multiple access channel is the dual of the broadcast erasure channel in the sense that the cut-set bound of the two channels are identical. This channel can be realized physically by using a computation code on the wireless channel, which computes the required linear combination. If all the channel coefficients in the wireless channel are good, then the combination $\sum_{i=1}^{d} X_i$ can be computed easily. However if one of the channel coefficients, say h_j, is in deep fade, then it may be algorithmically hard to compute this linear combination. Therefore, one way around this problem is to avoid having X_j in the linear combination and instead compute $\sum_{i=1,2,...,d\ i\neq j} X_i$. This gives rise to the channel model in (B.17).

The capacity of this multiple access channel is given by the cut-set bound (which is the same as the broadcast channel cut-set bound),

$$\mathcal{R}^{\text{MAC}} = \mathcal{R}^{\text{MAC}}_{\text{cut}} = \{(R_1, ..., R_D) | \sum_{i \in J} R_i \leq 1 - \epsilon^{|J|} \quad \forall J\}. \tag{B.18}$$

Here $J \subseteq \{1, 2, .., d\}$. We emphasize that the capacity region of this multiple access channel is equal to the cut-set bound of the erasure broadcast channel.

B.5 Proof of Lemma 24

We will first consider the case of a network with l transmit antennas and m single-antenna receivers. We can assume $l \leq m$, since if $l > m$, we can restrict ourselves to using m transmit antennas, which leaves the cut unaltered. Therefore $p = l$.

We can choose any particular subset of l receivers and use the strategy in Lemma 23 to achieve a DOF of $\frac{1}{\mathcal{O}(\log l)}$ for each receiver. We can time-share between all possible subsets of size l to achieve a certain DOF region. To compute the rate region achievable by this method, we use the following trick: Let the DOF tuple achieved be $\frac{1}{\mathcal{O}(\log l)}(r_1, ..., r_m)$. Let us construct a bi-partite graph with l nodes on the left partition and m nodes on the right partition and a complete graph connecting them. Each matching is equivalent to choosing a certain subset of the receivers (given by the set of right-partition nodes covered by the matching) and achieving DOF 1 for the each of the receivers. The characteristic vector of a bipartite matching is given by $(x_{ij})_M$ such that $x_{ij} = 1$ for edge (i, j) in the matching M and $x_{ij} = 0$ otherwise. The convex hull of these characteristic vectors is given by

$$\mathcal{M} = \text{conv}\left\{(x_{ij})_M | M \text{ a matching }\right\}. \tag{B.19}$$

For a bipartite graph, this is equivalent to the following polytope [131]

$$\mathcal{P} = \left\{ (x_{ij}) | x_{ij} \geq 0 \quad \forall i, j, \ \sum_j x_{ij} \leq 1 \quad \forall i \ \sum_i x_{ij} \leq 1 \quad \forall j \right\}. \tag{B.20}$$

The DOF d_j is given by

$$d_j = \frac{1}{\mathcal{O}(\log l)} \sum_i x_{ij}. \tag{B.21}$$

Now consider the following polytope,

$$\mathcal{D}_{\text{ach}} = \left\{ (d_j) | d_j \geq 0 \quad \forall j, \ d_j \leq \frac{1}{\mathcal{O}(\log l)}, \ \sum_j d_j \leq \frac{l}{\mathcal{O}(\log l)} \right\}. \tag{B.22}$$

We can show that $\mathcal{D}_{\text{ach}}$ is equivalent to $\mathcal{P}$ by using the mapping

$$\psi : \mathcal{D}_{\text{ach}} \rightarrow \mathcal{P} \tag{B.23}$$

$$\psi\{(d_j)\} = (x_{ij}) : x_{ij} = \mathcal{O}(\log l) \frac{d_j}{l}, \tag{B.24}$$

and the mapping

$$\zeta : \mathcal{P} \rightarrow \mathcal{D}_{\text{ach}} \tag{B.25}$$

$$\zeta\{(x_{ij})\} = (d_j) : d_j = \frac{1}{\mathcal{O}(\log l)} \sum_i x_{ij}. \tag{B.26}$$

Thus the region $\mathcal{D}_{\text{ach}}$ is achievable. Also the cut-set bound is given by

$$\mathcal{D}_{\text{cut}} = \left\{ (d_j) | d_j \geq 0 \quad \forall j, \ d_j \leq 1, \ \sum_j d_j \leq l \right\}. \tag{B.27}$$

This implies that

$$\mathcal{D}_{\text{ach}} = \frac{\mathcal{D}_{\text{cut}}}{\mathcal{O}(\log l)}, \tag{B.28}$$

which completes the proof of the single antenna receiver case.

For the multi-antenna receiver case, we will treat it as being composed of many single-antenna receivers, each receiving independent information and then sum up the rates. The proof will extend to this case to get the desired result.

B.6 Proof of Lemma 29

Let us start with a bi-partition cut on the polymatroidal network. This specifies a vertex partition Ω, which implicitly specifies the set of edges going between Ω and Ω^c as the edge cut, and a method to group certain edges together for charging the submodular constraints thus specifying the value of the cut. For example, consider the cut in Fig. 3.4b; it features the vertex partition and also specifies how to group edges to get an upper bound.

From the given partition we need to construct a cut on the Gaussian network. The polymatroidal cuts specify which of the edges need to be grouped together. By bounding the polymatroidal network in this manner, each edge is involved in *either* a broadcast constraint or a superposition constraint. This is in contrast to the Gaussian case, where each cut has a certain value, and there is no sense in which edges are assigned to broadcast or superposition constraint.

The key idea to connect these two cuts is the idea of *decoupling* the constraints in the Gaussian channel:

- In the Gaussian channel, if broadcast constraint is not active for a given edge, then the edge only participates in the superposition constraint and vice versa.
- While this is no longer true in the Gaussian network, we can obtain an upper bound network for the Gaussian network where this is true. The upper bound network is obtained by deactivating certain broadcast and superposition constraints. We refer to this process of obtaining an upper bound network where a certain constraint is not active as decoupling.
- Decoupling a *broadcast constraint* is easy because a network in which edges are not involved in a broadcast constraint can only, in general, do better than a network where there is a broadcast constraint on the edges.
- Decoupling the *superposition constraint* requires a bit more work; this can be illustrated using the following example. Suppose two edges e_1 and e_2 are involved in the superposition constraint in the following manner:

$$y = x_1 + x_2 + z, \tag{B.29}$$

with z being standard Gaussian noise. Then we construct another channel in which the edges e_1 and e_2 are not involved in a superposition constraint, i.e., the received symbol y comprises of two components y_1 and y_2 given by

$$y_i = x_i + z_i, i = 1, 2, \tag{B.30}$$

where z_i are independent Gaussian noise. This channel can emulate any scheme in the original channel if $\text{Var}(z_i) = \frac{\text{Var}(z)}{2}$. If this condition is satisfied, then we can add up y_1+y_2 to get $x_1+x_2+(z_1+z_2)$, which is statistically equivalent to the original channel. Therefore, to decouple the superposition constraint involving d variables, we need to reduce the variance of the noise by d, the degree of the superposition constraint, or equivalently, increase the signal power by a factor d.

- Therefore we can decouple all required broadcast and superposition constraints, if the power is increased by a factor of d, which is the maximum degree of any node.

Thus given any cut in the polymatroidal network along with the assignments of the edges to broadcast or superposition constraints, we can obtain a similar cut in the Gaussian network by decoupling the constraints which are not active in the polymatroidal cut. This incurs a power penalty factor of d, thus as far as the outer bound is concerned, we can assume that each node has power dP instead of P. The network thus obtained is made of MAC and broadcast channels. In this network, every cut decomposes into the sum of MAC and BC cuts. A MAC cut with d nodes when evaluated under general distribution on the input, is of the form

$$\begin{aligned}
\sum_i R_{ij} &\leq \mathbb{E}\log\left(1+(\sum_i |h_{ij}|)^2 dP\right) && \text{(B.31)}\\
&\leq \mathbb{E}\log\left(1+\sum_i |h_{ij}|^2 d^2 P\right) && \text{by Cauchy-Schwartz inequality}\\
&\leq \log\left(1+\mathbb{E}(\sum_i |h_{ij}|^2) d^2 P\right) && \text{by Jensen inequality}\\
&= \log\left(1+d^3P\right) && \text{(B.32)}\\
&\leq \mathbb{E}\left\{\log\left(1+ad^3P|h|^2\right)\right\} = C(ad^3P). && \text{(B.33)}
\end{aligned}$$

The last step follows because,

$$\mathbb{E}\left\{\log\left(1+c|h|^2\right)\right\} = \mathbb{E}\left\{\log\left(1+ce^{\log|h|^2}\right)\right\} \tag{B.34}$$

$$\geq \log\left(1+ce^{\mathbb{E}(\log|h|^2)}\right). \tag{B.35}$$

Here $a := e^{-\mathbb{E}(\log|h|^2)}$ is finite for the fading distribution by assumption. Thus, the cutset bound for the original network implies that

$$\sum_i R_{ij} \leq C\left(ad^3P\right), \tag{B.36}$$

whereas the corresponding cutset bound for the polymatroidal network is of the form

$$\sum_i R_{ij} \leq \frac{1}{2}C(2P). \tag{B.37}$$

We get

$$\mathcal{R}_{\text{cut}}^{\text{poly}} \supseteq \frac{1}{2}\mathcal{R}_{\text{cut}}^{\text{original}}\left(\frac{P}{bd^3}\right), \tag{B.38}$$

where $b := \frac{a}{2} = \frac{e^{\mathbb{E}(\log|h|^2)}}{2}$ is a constant depending on the fading distribution. For h distributed as complex Gaussian, $b \approx 0.86$.

B.7 Proof of Theorem 24

The k sources are in the layer V_0, and the k destinations are in the layer V_{L+1}. Let the number of nodes in the i-th layer be n_i.

Max-Flow Rate

Let R_i be the rate between the i-th source destination pair. We will route the flow in a symmetric manner, where the incoming flow is divided equally among all the edges going out of a node. We will compute constraints on the rate region achievable by this strategy.

In the first hop, all edges going out of the i-th source will carry a flow of value

$\frac{R_i}{n_1}$. The constraint imposed by the edges going out of the source is given by

$$\frac{R_i}{n_1} n_1 \leq 1 \iff R_i \leq 1, \quad \forall i = 1, 2, .., k. \tag{B.39}$$

The constraint imposed by the edges coming into the nodes of the first hop are given by

$$\frac{\sum_i R_i}{k n_1} k \leq 1 \iff \sum_i R_i \leq n_1. \tag{B.40}$$

The total flow carried by all the nodes in any given layer equals $\sum_i R_i$, and each node in layer l carries a flow of $\frac{R_i}{n_l}$ corresponding to flow i.

In the $l-th$ hop connecting layers V_{l-1} and V_l, each edge carries a flow of value $\frac{R_i}{n_l n_{l+1}}$ corresponding to source i, which yields a total flow of value $\frac{\sum_i R_i}{n_l n_{l+1}}$ for each edge. The outgoing constraints on layer $l-1$ yield

$$\frac{\sum_i R_i}{n_l n_{l+1}} n_{l+1} \leq 1 \iff \sum_i R_i \leq n_l, \quad \forall i = 1, 2, .., k. \tag{B.41}$$

The incoming constraints on layer l yield

$$\frac{\sum_i R_i}{n_l n_{l+1}} n_l \leq 1 \iff \sum_i R_i \leq n_{l+1}, \quad \forall i = 1, 2, .., k. \tag{B.42}$$

In the final (L+1-th) hop also, there are constraints similar to layer 1. In particular, the outgoing constraints on layer l yield

$$\frac{\sum_i R_i}{n_L k} k \leq 1 \iff \sum_i R_L \leq n_L. \tag{B.43}$$

and the incoming constraints on the destination layer yield

$$\frac{R_i}{n_L} n_L \leq 1 \iff R_i \leq 1, \quad \forall i = 1, 2, .., k. \tag{B.44}$$

Thus a rate pair $(R_1, ..., R_k)$ is achievable by routing iff

$$\sum_i R_i \leq \min(n_l, n_{l+1}) \quad \forall l = 0, 1, ..., L. \tag{B.45}$$

$$R_i \leq 1 \quad \forall i = 1, 2, ..., k. \tag{B.46}$$

Cut-set region We can easily write the following constraints, which are a subset of the cut-set bounds.

Corresponding to the cut separating $\Omega = \cup_{i=0}^{l} V_i$ and Ω^c, the following constraint can be written,

$$\sum_i R_i \leq \min(n_l, n_{l+1}) \quad \forall l = 0, 1, ..., L. \tag{B.47}$$

Corresponding to the cut given by $\Omega = s_i$, we can get the following constraint

$$R_i \leq 1 \quad \forall i = 1, ..., k. \tag{B.48}$$

Comparing (B.45), (B.46) and (B.47), (B.48), we can deduce that any rate tuple that satisfies the cut-set region will lie in the rate region achieved by the flow. Thus the rate region corresponding to max-flow equals the rate region corresponding to cut-set region.

For example, Fig. 3.5b denotes the directed layered polymatroidal network obtained from the network in Fig. 3.5a. Every node basically constrains the total inflow and outflow to be lesser than 1.

REFERENCES

[1] S. Aeron and V. Saligrama, "Wireless ad hoc networks: Strategies and scaling laws for the fixed SNR regime," *IEEE Trans. Inform. Theory,* vol. 53, no. 6, pp. 2044–2059, Jun. 2007.

[2] A. Agarwal, N. Alon and M. Charikar, "Improved approximation for directed cut problems," in *Proc. of ACM STOC*, 2007, pp. 671–680.

[3] A. Agarwal and M. Charikar, "On the advantage of Network Coding for Improving Network Throughput," in *Proc. Information Theory Workshop*, 2004, pp. 247–249.

[4] R. Ahlswede, C. Ning, S. Li, and R. Yeung, "Network information flow," *IEEE Trans. Inform. Theory,* vol. 46, no. 4, pp. 1204-1216, July 2000.

[5] R. Ahlswede and Ning Cai, "On communication complexity of vector-valued functions," *IEEE Trans. Inform. Theory*, vol. 40, no.6, pp. 2062–2067, 1994.

[6] D. Aldous, C. McDiarmid, and A. Scott, "Uniform Multicommodity Flow through the Complete Graph with Random Edge-capacities," *Operations Res. Letters,* vol. 37, pp. 299-302, 2009.

[7] M. A. M. Ali and D. Tse, "Completely Stale Transmitter Channel State Information is Still Very Useful," presented at Forty-eigth Annual Allerton Conference, Monticello, IL, Oct. 2010.

[8] A. Amadruz, and C. Fragouli, "Combinatorial algorithms for wireless information flow," in *Proc. of ACM-SIAM SODA*, Jan. 2009, pp. 555–564.

[9] M. Anand and P. R. Kumar, "A Digital Interface for Gaussian Relay and Interference Networks: Lifting Codes From the Discrete Superposition Model," *IEEE Trans. Inform. Theory,* vol. 57, no. 5, pp. 2548–2564.

[10] A. Anandkumar, M. Wang, L. Tong, and A. Swami, "Prize-Collecting Data Fusion for Cost-Performance Tradeoff in Distributed Inference," in *Proc. of IEEE INFOCOM,* 2009, pp. 2150–2158 .

[11] V. S. Annapureddy, A. El Gamal, and V. V. Veeravalli, "Degrees of Freedom of Interference Channels with CoMP Transmission and Reception," Sep. 2011. [Online]. Available: http://arxiv.org/abs/1109.5396

[12] R. Appuswamy, M. Franceschetti, N. Karamchandani, and K. Zeger, "Network coding for computing: Cut-set bounds," *IEEE Trans. Inform. Theory*, vol. 57, no.2, pp. 1015–1030, Feb. 2011.

[13] R. Appuswamy, M. Franceschetti, N. Karamchandani, and K. Zeger, "Linear Codes, Target Function Classes, and Network Computing Capacity," Feb. 2011. [Online]. Available: http://arxiv.org/abs/1102.4825

[14] S. Arora, S. Rao and U. Vazirani, "Expander Flows, Geometric Embeddings, and Graph Partitionings," *Journal of the ACM*, vol.56, no.2, pp. 1–37, 2009. Preliminary version in *Proc. of ACM STOC*, 2004, pp. 222–231.

[15] S. Arora, J. Lee and A. Naor, "Euclidean distortion and the Sparsest Cut," *J. of the AMS*, vol. 21, no. 1, pp. 1–27, 2008. Preliminary version in *Proc. of ACM STOC*, 2005, pp. 553-562.

[16] Y. Aumann and Y. Rabani, "An $O(\log k)$ approximate min-cut max-flow theorem and approximation algorithm," *SIAM Journal on Computing*, vol. 27, no. 1, pp. 291–301, 1998.

[17] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse, "Wireless Network Information Flow: A Deterministic Approach," *IEEE Trans. Inform. Theory,* vol. 57, no. 4, pp. 1872–1905, 2011.

[18] O. Ayaso, D. Shah, and M. Dahleh, "Lower bounds on information rates for distributed computation via noisy channels," presented at Forty-fifth Annual Allerton Conference, Monticello, IL, 2007.

[19] J. Bourgain, "On Lipschitz embedding of finite metric spaces in Hilbert space," *Israeli J. Math.*, vol. 52, no. 1, pp. 46–52, 1985.

[20] B. Brinkman, A. Karagiozova, and J. Lee, "Vertex cuts, random walks, and dimension reduction in series-parallel graphs," in *Proc. of ACM STOC*, 2007, pp.621–630.

[21] V. Cadambe and S. Jafar, "Interference alignment and the degrees of freedom of the K user interference channel," *IEEE Trans. Inform. Theory,* vol. 54, no. 8, pp. 3425-3441, Aug. 2008.

[22] T. H. Chan and A. Grant, "Mission impossible: Computing the network coding capacity region," in *Proc. IEEE Int. Symp. Inform. Theory,* Jul. 2008, pp. 320–324.

[23] M. Charikar, T. Cheung, Z. Dai, A. Goel, S. Guha, and M. Li, "Approximation Algorithms for Directed Steiner Problems," *Journal of Algorithms*, vol. 33, pp. 73–91, 1999.

[24] C. Chekuri, "Approximation Algorithms," lecture notes for CS 598CSC, Department of Computer Science, University of Illinois, Urbana-Champaign, 2011. [Online]. Available: `http://www.cs.illinois.edu/class/sp11/cs598csc/`

[25] C. Chekuri, S. Kannan, A. Raja and P. Viswanath, "Multicommodity Flows and Cuts in Polymatroidal Networks," in *Proc. Innovations in Theoretical Computer Science,* 2012, pp. 399–408. [Online]. Available: `http://arxiv.org/abs/1110.6832`

[26] J. Cheriyan, H. Karloff and Y. Rabani, "Approximating Directed Multicuts," *Combinatorica*, vol. 25, no. 3, pp. 251–269, 2005.

[27] M. Chiang, S. H. Low, A. R. Calderbank, and J. C. Doyle, "Layering as Optimization Decomposition: A Mathematical Theory of Network Architectures," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 255–312, Jan. 2007.

[28] J. Chuzhoy and S. Khanna, "Polynomial Flow-Cut Gaps and Hardness of Directed Cut Problems," *Journal of the ACM*, vol. 56, no. 2, 2009. Preliminary version in *Proc. of the ACM STOC*, 2007, pp. 179–188.

[29] "Cisco Global Mobile Data Forecast," Cisco White paper, Feb. 2012. [Online]. Available: `http://goo.gl/aYXM3`

[30] T. H. Cormen, C. E. Leiserson, R. L. Rivest and C. Stein, *Introduction to algorithms*, MIT Press, 2001.

[31] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed., John Wiley and Sons, New York, 2006.

[32] A. F. Dana, M. Sharif, R. Gowaikar, B. Hassibi, and M. Effros, "Is broadcast plus multiaccess optimal for Gaussian wireless networks?" in *Proc. Asilomar Conf. on Signals, Systems and Computers,* 2003, pp. 1748—1752.

[33] A. F. Dana, R. Gowaikar, R. Palanki, B. Hassibi, and M. Effros, "Capacity of wireless erasure networks," *IEEE Trans. Inform. Theory,* vol. 52, no. 3, pp. 789-804, March 2006.

[34] V. Doshi, D. Shah, and M. Medard, "Source coding with distortion through graph coloring," in *Proc. IEEE Int. Symp. Inform. Theory,* 2007, pp. 1501–1505.

[35] C. Dutta, Y. Kanoria, D. Manjunath, and J. Radhakrishnan, "A tight lower bound for parity in noisy communication networks," in *Proc. of ACM-SIAM SODA*, pp. 1056–1065, January 2008.

[36] J. Edmonds, "Submodular functions, matroids and certain polyhedra," in *Proc. Calgary Int. Conf. Combinatorial Structures and Applications.* pp. 69–87, June 1969.

[37] J. Edmonds and E. Giles, "A min-max relation for submodular functions on graphs," *Annals of Discrete Mathematics*, vol.1, no. 1, 185–204, 1977.

[38] A. El Gamal, "On Information Flow in Relay Networks," in *Proc. IEEE National Telecomm. Conf.,* vol. 2, pp. D4.1.1–D4.1.4, Nov. 1981.

[39] R. Etkin and E. Ordentlich, "On the degrees-of-freedom of the K-user Gaussian interference channel," Jan. 2009. [Online]. Available: `http://arxiv.org/abs/0901.1695`

[40] R. Etkin, D. Tse, and H. Wang, "Gaussian Interference Channel Capacity to Within One Bit," *IEEE Trans. Inform. Theory*, vol. 54, no. 12, pp. 5534–5562, Dec. 2008.

[41] G. Even, J. Naor, S. Rao and B. Schieber, "Divide-and-conquer approximation algorithms via spreading metrics," in *Journal of the ACM*, vol. 47, no. 4, pp. 585–616, 2000.

[42] J. Fackaroenphol, S. Rao and K. Talwar, "A tight bound on approximating arbitrary metrics by tree metrics," *Journal of Computer and System Sciences* vol. 69, no. 3, pp.485–497, 2004.

[43] J. Fakcharoenphol and K. Talwar, "An Improved Decomposition Theorem for Graphs Excluding a Fixed Minor," in *Proc. Random-Approx*, Springer Verlag, 2003, pp. 36-46.

[44] A. Federgruen and H. Groenevelt, "Polymatroidal flow network models with multiple sinks," *Networks,*, vol.18, no. 4, pp. 285–302, 1988.

[45] U. Feige, M.T. Hajiaghayi, and J. Lee, "Improved approximation algorithms for minimum-weight vertex separators," *SIAM J. on Computing*, vol. 38, no. 2, pp. 629–657, 2008. Preliminary version in *Proc. of ACM STOC*, 2005, pp. 563–572.

[46] M. Franceschetti, M. D. Migliore, P.. Minero, and F. Schettino, "The Degrees of Freedom of Wireless Networks Via Cut-Set Integrals," *IEEE Trans. Inform. Theory,* vol. 57, no. 5, pp. 3067–3079.

[47] A. Frank, "Edge-connection of Graphs, Digraphs, and Hypergraphs," Tech. Rep., Enervary Research Group on Combinatorial Optimization, Budapest, Hungary, Sep. 2000. [Online]. Available: `http://www.cs.elte.hu/egres`.

[48] N. Garg, V. Vazirani, and M. Yannakakis, "Approximate Max-Flow Min-(Multi)Cut Theorems and Their Applications," *SIAM J. Comput.*, vol. 25, no. 2, pp. 235–251, 1996.

[49] M. Gatzianas, L. Georgiadis, and L. Tassiulas, "Multiuser broadcast erasure channel with feedback–capacity and algorithms," Sep. 2010. [Online]. Available: `http://arxiv.org/abs/1009.1254`

[50] L. Georgiadis and L. Tassiulas, "Broadcast erasure channel with feedback capacity and algorithms," in *Proc. 5th Workshop on Net. Cod. Theory and Appl.,* June 2009, pp. 54-61.

[51] A. Giridhar and P. R. Kumar, "Computing and communicating functions over sensor networks," *IEEE Journal on Selected Areas in Communication*, vol. 23, no. 4, pp. 755–764, Apr. 2005.

[52] M. X. Goemans, S. Iwata, and R. Zenklusen, "A flow model based on polylinking system," *Math. Programming Ser A*, vol. 135, no. 1, 2012.

[53] T. Gou, S. A. Jafar, S.-W. Jeon, and S.-Y. Chung, "Aligned Interference Neutralization and the Degrees of Freedom of the $2 \times 2 \times 2$ Interference Channel," Dec. 2010. [Online]. Available: `http://arxiv.org/abs/1012.2350v1`

[54] A. Gupta, "Improved results for directed multicut," in *Proc. of ACM-SIAM SODA*, 2003, pp. 454–455.

[55] P. Gupta and P. R. Kumar, "Capacity of Wireless Networks," *IEEE Trans. Inform. Theory,* vol. 46, no. 2, pp. 388-404, March 2000.

[56] A. Gupta, V. Nagarajan and R. Ravi, "An improved approximation algorithms for requirement cut," *Operations Research Letters* vol. 38, no. 4, pp.322-325, 2010.

[57] A. Gupta, I. Newman, Y. Rabinovich, and A. Sinclair, "Cuts, trees and ℓ_1-embeddings of graphs," *Combinatorica*, vol. 24, pp. 233–269, 2004. Preliminary version in *Proc. of IEEE FOCS*, 1999, p. 399.

[58] D. Gunduz, A. Yener, A. Goldsmith, and H. V. Poor, "The Multi-way Relay Channel," Apr. 2010 [Online]. Available: `http://arxiv.org/abs/1004.2434`

[59] N. J. A. Harvey, R. Kleinberg, and A. R. Lehman, "On the capacity of information networks," *IEEE Trans. Inform. Theory,* vol. 52, no.6, pp. 241-250, 2006.

[60] R. Hassin, "On Network Flows," Ph.D Dissertation, Yale University, 1978.

[61] Syed A. Jafar, "Interference Alignment: A New Look at Signal Dimensions in a Communication Network," *Foundations and Trends in Communications and Information Theory*, vol. 7, no. 1, pp. 1–136, 2011.

[62] S. Jafar and S. Shamai, "Degrees of Freedom Region of the MIMO X Channel," *IEEE Trans. Inform. Theory,* vol. 54, no. 1, pp. 151-170, Jan. 2008.

[63] S. -W. Jeon, S. -Y. Chung, and S. Jafar, "Degrees of freedom of multi-source wireless relay networks," presented at Forty-seventh Annual Allerton Conference, Monticello, IL, Sep. 2009.

[64] S. -W. Jeon; S. -Y. Chung; S. A. Jafar, "Degrees of Freedom Region of a Class of Multisource Gaussian Relay Networks," *IEEE Trans. Inform. Theory,* vol. 57, no. 5, pp. 3032–3044.

[65] N. Kahale, "On reducing the cut ratio to the multicut problem," Technical Report TR-93-78, DIMACS 1993.

[66] S. Kamath, V. Ananthram and D. Tse, "Generalized Network Sharing Outer Bound and the Two-Unicast Problem," in *Proc. 7th Workshop on Net. Cod. Theory and Appl.,* July 2011, pp. 1–6.

[67] S. Kamath, S. Kannan and P. Viswanath, "Wireless Networks with Symmetric Demands," in *Proc. IEEE Int. Symp. Inform. Theory,* July 2012, pp. 2621 – 2625.

[68] S. Kamath and P. Viswanath, "An information-theoretic meta-theorem on edge-cut bounds," in *Proc. IEEE Int. Symp. Inform. Theory,* July 2012, pp. 1657 – 1661.

[69] S. Kannan, A. Raja and P. Viswanath, "Approximately Optimal Broadcasting-cum-Multicasting in Wireless Networks," in *Proc. IEEE Int. Symp. Inform. Theory,* Aug. 2011, pp. 1911–1915.

[70] S. Kannan, A. Raja and P. Viswanath, "Approximately Optimal Broadcasting in Wireless Networks," *IEEE Trans. Inform. Theory*, to be published.

[71] S. Kannan, A. Raja and P. Viswanath, "Local Phy + Global Flow: A Layering Principle for Wireless Networks," in *Proc. IEEE Int. Symp. Inform. Theory,* Aug. 2011, pp. 1633–1637,.

[72] S. Kannan and P. Viswanath, "Multiple-Unicast in Fading Wireless Networks: A Separation Scheme is Approximately Optimal," in *Proc. IEEE Int. Symp. Inform. Theory,* Aug. 2011, pp. 2617–2621.

[73] S. Kannan and P. Viswanath, "Approximate capacity theorems for multiple unicast in wireless networks," *IEEE Trans. Inform. Theory*, submitted.

[74] S. Kannan and P. Viswanath, "Multi-session function computation in undirected graphs," May 2011. [Online]. Available: http://www.ifp.illinois.edu/~pramodv/pubs/KV12_2.pdf

[75] M. Kim, E. Erez, E. M. Yeh, and M. Medard, "Deterministic Network Model Revisited: An Algebraic Network Coding Approach," Mar. 2011. [Online]. Available: http://arxiv.org/abs/1103.0999

[76] P. Klein, S. Plotkin and S. Rao, "Planar graphs, multicommodity flow, and network decomposition," in *Proc. of ACM STOC*, 1993, pp. 682–690.

[77] P. N. Klein, S. A. Plotkin, S. Rao, and E. Tardos, "Approximation Algorithms for Steiner and Directed Multicuts," *J. Algorithms*, vol. 22, no. 2, pp. 241–269, 1997.

[78] G. Kramer and S. A. Savari, "Edge-cut bounds on network coding rates," *J. Network and Systems Management,* vol. 14, no. 1, pp. 49-67, March 2006.

[79] R. Koetter and M. Medard, "An algebraic approach to network coding," *IEEE Trans. Networking,* vol.11, no.5, pp. 782–795, Oct. 2003.

[80] R. Koetter, M. Effros, and M. Medard, "A Theory of Network Equivalence – Part I: Point to Point Channels," *IEEE Trans. Inform. Theory,* vol. 57, no. 2, pp. 972–995, Feb. 2011.

[81] R. Koetter, M. Effros, and M. Medard, "A Theory of Network Equivalence – Part II: Multiterminal Channels," Jul. 2010. [Online]. Available: http://arxiv.org/abs/1007.1033.

[82] H. Kowshik and P. R. Kumar, "Optimal Function Computation in Directed and Undirected Graphs," May 2011. [Online]. Available: http://arxiv.org/abs/1105.0240.

[83] W.-C. Kua and C.-C. Wang, "On the Capacity of 2-User 1-Hop Relay Erasure Networks the Union of Feedback, Scheduling, Opportunistic Routing, and Network Coding," in *Proc. IEEE Int. Symp. Inform. Theory,* Aug. 2011, pp. 1337–1341.

[84] K. Kuratowski, "Sue le probléme des courbes gauches en topologie," *Fund. Math.*, vol. 15, pp. 217–283, 1930.

[85] E. Kushilevitz and N. Nisan, *Communication Complexity*, Cambridge University Press, 1997.

[86] M. Langberg and M. Medard, "The Multiple Unicast Network Coding Conjecture," presented at Forty-seventh Annual Allerton Conference, Monticello, IL, Oct. 2009.

[87] M. Langberg and A. Ramamoorthy, "Communicating the sum of sources in a 3-sources/3-terminals network," in *Proc. IEEE Int. Symp. Inform. Theory,* 2009, pp. 2121–2125.

[88] E. L. Lawler and C. U. Martel, "Computing maximal Polymatroidal network flows," *Math. Oper. Res.,*, vol. 7, no. 3, pp. 334–347, 1982.

[89] J. Lee and C. Rashtchian, "A Simpler Proof Of The KPR Theorem," Jan. 2012. [Online]. Available: `http://tcsmath.wordpress.com/2012/01/11/a-simpler-proof-of-the-kpr-theorem/`

[90] F. T. Leighton and S. Rao, "Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms," *Journal of the ACM*, vol. 46, no. 6, pp. 787–832, 1999. Preliminary version in *Proc. of ACM STOC*, 1988.

[91] S. H. Lim; Y. -H. Kim; A.. El Gamal, and S. -Y. Chung, "Noisy Network Coding," *IEEE Trans. Inform. Theory,* vol. 57, no. 5, pp. 3132–3152, 2011.

[92] Z. Li and B. Li, "Network Coding: The Case of Multiple Unicast Sessions," presented at Forty-second Annual Allerton Conference, Monticello, IL, 2004.

[93] Z. Li and B. Li, "Network Coding in Undirected Networks," presented at Thirty-eigth Annual Conference on Information Sciences and Systems, Princeton, NJ, Mar., 2004.

[94] D. S. Lun, M. Medard, R. Koetter, and M. Effros, "On coding for reliable communication over packet networks," *Physical Communication*, vol. 1, no. 1, pp. 3–20, 2008.

[95] A. Orlitsky and J. R. Roche, "Coding for computing," *IEEE Trans. Inform. Theory,* 47, no. 3, pp. 903–917, 2001.

[96] O. Leveque and I. E. Telatar, "Information-theoretic upper bounds on the capacity of large, extended ad hoc wireless networks. *IEEE Trans. Inform. Theory,* vol. 51, no. 3, pp. 858–865, Mar. 2005.

[97] N. Linial, E. London, and Y. Rabinovich, "The Geometry of Graphs and Some of its Algorithmic Applications," *Combinatorica*, vol. 15, no. 2, pp. 215–245, 1995.

[98] L. Lovász, "Submodular functions and convexity," in *Mathematical programming: the state of the art*, Springer Verlag, 1983, pp. 235–257.

[99] N. Ma, P. Ishwar, and P. Gupta, "Information-theoretic bounds for multiround function computation in collocated networks," in *Proc. IEEE Int. Symp. Inform. Theory,* 2009, pp. 2306–2310.

[100] M. A. Maddah-Ali, S. A. Motahari, and Amir K. Khandani, "Communication over MIMO X Channels: Signaling and Performance Analysis," *IEEE Trans. Inform. Theory*, vol. 54, no. 8, pp. 3457–3470, 2008.

[101] I. Maric, A. Goldsmith, and M. Medard, "Multihop Analog Network Coding via Amplify-and-Forward: The High SNR Regime," *IEEE Trans. Inform. Theory*, vol. 58, no. 2, pp. 793 - 803, Feb. 2012.

[102] C. Martel, "Preemptive Scheduling with Release Times, Deadlines, and Due Times," *Journal of the ACM*, vol. 29, no. 3, pp. 812–829, 1982.

[103] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inform. Theory,* vol. 25, no. 3, pp. 306–311, 1979.

[104] J. Matousek and Y. Rabinovich, "On dominated ℓ_1 metric," *Israeli J. of Mathematics*, vol. 123, no. 1, pp. 285–301, 2001.

[105] S. Mohajer, S. N. Diggavi, C. Fragouli, and D. N. C. Tse, "Approximate Capacity of Gaussian Interference-Relay Networks," *IEEE Trans. Inform. Theory,* vol. 57, no. 5, pp. 2837–2684, 2011.

[106] D. Mosk-Aoyama and D. Shah, "Fast distributed algorithms for computing separable functions," *IEEE Trans. Inform. Theory,* vol. 54, no. 7, pp. 2997–3007, Jul. 2008.

[107] A. Motahari, S. O. Gharan, and A. Khandani, "Real interference alignment with real numbers," Aug. 2009. [Online]. Available: `http://arxiv.org/abs/0908.1208`

[108] V. Nagarajan and R. Ravi, "Approximation Algorithms for requirement cut on graphs," *Algorithmica* vol. 56, no. 2, pp. 198–213, 2010.

[109] J. Naor and L. Zosin, "A 2-approximation algorithm for the directed multiway cut problem," *SIAM J. on Computing*, vol. 31, no. 2, pp. 477–482, 2001. Preliminary version in *Proc. of IEEE FOCS*, 1997.

[110] C. St. J. A. Nash-Williams, "Edge-disjoint Spanning Trees of Finite Graphs," *J. London Math. Soc.,* vol. 36, pp. 445450, 1961.

[111] B. Nazer and M. Gastpar, "Computing over multiple-access channels," *IEEE Trans. Inform. Theory,* vol. 53, no. 10, pp. 3498–3516, Oct. 2007.

[112] B. Nazer, M. Gastpar, S. Jafar, and S. Vishwanath, "Ergodic interference alignment," in *Proc. IEEE Int. Symp. Inform. Theory,* Seoul, 2009.

[113] D. O'Neill, B.S. Thian, A. Goldsmith and S. Boyd, "Wireless NUM: Rate and Reliability Tradeoffs in Random Environments," in *Proc. IEEE WCNC*, Budapest Hungary, 2009.

[114] U. Niesen, "Interference Alignment in Dense Wireless Networks," Dec. 2009. [Online]. Available: `http://arxiv.org/abs/0912.0868`

[115] U. Niesen, P. Gupta, D. Shah, "The Balanced Unicast and Multicast Capacity Regions of Large Wireless Networks," *IEEE Trans. Inform. Theory,* vol. 56, no. 5, pp. 2249-2271, 2010.

[116] U. Niesen, P. Gupta, D. Shah, "On Capacity Scaling in Arbitrary Wireless Networks," *IEEE Trans. Inform. Theory,* vol. 55, no. 9, pp. 3959–3982, 2009.

[117] U. Niesen, B. Nazer and P. Whitting, "Computation Alignment: Capacity Approximation without Noise Accumulation," Aug. 2011. [Online]. Available: `http://arxiv.org/abs/1108.6312`

[118] J. Oxley, *Matroid Theory*, Oxford University Press, 2006.

[119] O. Oyman, R. U. Nabar, H. Bolcskei, and A. J. Paulraj, "Tight lower bounds on the ergodic capacity of Rayleigh fading MIMO channels," in *Proc. IEEE Globecom*, vol. 2 , 2002, pp.1172 – 1176 .

[120] A. Ozgur, R. Johari, D. Tse, and O. Leveque, "Information Theoretic Operating Regimes of Large Wireless Networks," *IEEE Trans. Inform. Theory,* vol.56, no.1, pp. 427–437, 2010.

[121] A. Ozgur, O. Leveque and D. N. C. Tse, "Hierarchical Cooperation Achieves Optimal Capacity Scaling in Ad Hoc Networks," *IEEE Trans. Inform. Theory,* vol. 53, no. 10, pp. 3549–3572, October 2007.

[122] S. Plotkin and E. Tardos, "Improved Bounds on the Max-Flow Min-Cut Ratio for Multicommodity Flows," *Combinatorica*, vol. 15, no. 3, pp. 425–434, 1995.

[123] Y. Rabinovich. "On Average Distortion of Embedding Metrics into the Line," *Discrete & Computational Geometry*, vol. 39, no. 4, pp. 720–733, 2008. Preliminary version in *Proc. of ACM STOC*, 2003.

[124] A. Raja and P. Viswanath, "Compress-and-Forward Scheme for a Relay Network: Approximate Optimality and Connection to Algebraic Flows," in *Proc. IEEE Int. Symp. Inform. Theory,* Aug. 2011, pp.1698–1702.

[125] B. K. Rai and B. K. Dey, "Feasible alphabets for communicating the sum of sources over a network," in *Proc. IEEE Int. Symp. Inform. Theory,* 2009, pp. 1353–1357.

[126] B. K. Rai, and B. K. Dey, "Sum-networks: System of polynomial equations, unachievability of coding capacity, reversibility, insufficiency of linear network coding," Jun. 2009. [Online]. Available: `http://arxiv.org/abs/0906.0695`

[127] B. K. Rai, B. K. Dey, and S. Shenvi, "Some bounds on the capacity of communicating the sum of sources," in *Proc. ITW 2010, Cairo*, 2010, pp. 1–5.

[128] A. Ramamoorthy, "Communicating the sum of sources over a network," in *Proc. IEEE Int. Symp. Inform. Theory,* 2008, pp. 1646–1650.

[129] S. Rao, "Small Distortion and Volume Preserving Embeddings for Planar and Euclidean Metrics," *Proc. of ACM Symposium on Computational Geometry*, 1999, pp. 300–306.

[130] M. Saks, A. Samorodnitsky and L. Zosin, "A Lower Bound On The Integrality Gap For Minimum Multicut In Directed Networks," *Combinatorica*, vol. 24, no. 3, pp. 525–530, 2004.

[131] A. Schrijver, "Combinatorial Optimization: Polyhedra and Efficiency," Springer-Verlag, 2003.

[132] A. Schrijver, "Matroids and linking systems," *Journal of Combinatorial Theory B*, vol. 26, no. 3, pp. 349–369, 1979.

[133] V. Shah, B. Dey, and D. Manjunath, "Network Flows for Functions, " in *Proc. IEEE Int. Symp. Inform. Theory,* St. Petersburg, Russia, Aug. 2011, pp. 234–238.

[134] S. Shakkottai and R. Srikant, "Network Optimization and Control," *Foundations and Trends in Networking*, NoW Publishers, 2007.

[135] C. E. Shannon, "A Mathematical Theory of Communication," *Bell Syst. Tech. J.* vol. 27, pp. 379–423, 1948.

[136] D. Shmoys, "(Approximation algorithms for) Cut problems and their application to divide-and-conquer," *Approximation Algorithms for NP-hard Problems*, (D.S. Hochbaum, Ed.) PWS, 1997, 192-235.

[137] I. Shomorony and A. S. Avestimehr, "Two-Unicast Wireless Networks: Characterizing the Degrees-of-Freedom," Feb. 2011. [Online]. Available: `http://arxiv.org/abs/1102.2498`

[138] S. Subramanian, P. Gupta, and S. Shakkottai, "Scaling bounds for function computation over large networks," in *Proc. IEEE Int. Symp. Inform. Theory,* 2007, pp. 136–140.

[139] C. Suh and D. N. Tse, "Feedback Capacity of the Gaussian Interference Channel to Within 2 Bits," *IEEE Trans. Inform. Theory,* vol. 57, no. 5, pp. 2667–2685,2011.

[140] E. Tardos and V. Vazirani, "Improved Bounds for the Max-Flow Min-Multicut Ratio for Planar and $K_{r,r}$-Free Graphs," *Information Processing Letters*, vol. 47, no. 2, pp. 77–80, 1993.

[141] C. Tian, J. Chen, S. N. Diggavi, and S. Shamai, "Optimality and Approximate Optimality of Source-Channel Separation in Networks," in *Proc. IEEE Int. Symp. Inform. Theory,* 2010, pp. 495–499.

[142] D. Tse and S. Hanly, "Multi-Access Fading Channels: Part I: Polymatroid Structure, Optimal Resource Allocation and Throughput Capacities. *IEEE Trans. Inform. Theory,* vol. 44, no. 7, pp. 2796–2815, Nov. 1997.

[143] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication,* Cambridge University Press, New York, 2005.

[144] W. T. Tutte, "On the Problem of Decomposing a Graph Into n Connected Factors," *J. London Math. Soc.*, vol. 36, no. 1, pp. 221230, 1961.

[145] A. Vahid, V. Aggarwal, A. S. Avestimehr, and A. Sabharwal, "On the Capacity of Multi-Hop Wireless Networks with Partial Network Knowledge," Oct. 2010. [Online]. Available: `http://arxiv.org/abs/1010.1060`

[146] D. Vasudevan and S. Korada, "Polymatroidal Flows on Two Classes of Information Networks," *IEEE Trans. Inform. Theory,* vol. 57, no.1, pp. 227–233, 2011.

[147] C. S. Vaze and M. K. Varanasi, "The Degrees of Freedom of the 2-Hop, 2-User Interference Channel with Feedback," presented at Forty-ninth Annual Allerton Conference, Monticello, IL, 2011.

[148] S. Vishwanath, N. Jindal, and A. J. Goldsmith, "Duality, achievable rates and sum-rate capacity of Gaussian MIMO broadcast channel," *IEEE Trans. Inform. Theory,* vol. 49, no. 10, pp. 2658–2668, 2003.

[149] P. Viswanath and D. Tse, "Sum capacity of the multiple antenna Gaussian broadcast channel and uplink-downlink duality," *IEEE Trans. Inform. Theory,* vol. 49, no. 8, pp. 1912–1923, Aug. 2003.

[150] C.-C. Wang, "Capacity of 1-to–K broadcast packet erasure channels with channel output feedback," presented at Forty-eigth Annual Allerton Conference, Monticello, IL, 2010.

[151] I. Wang and D. N. Tse, "Interference Mitigation Through Limited Receiver Cooperation," *IEEE Trans. Inform. Theory,* vol. 57, no. 5, pp. 2913–2940.

[152] I. Wang and D. N. Tse, "Interference Mitigation Through Limited Transmitter Cooperation," *IEEE Trans. Inform. Theory,* vol. 57, no. 5, pp. 2941–2965.

[153] H. Weingarten, Y. Steinberg, S. Shamai, "The Capacity Region of the Gaussian MIMO Broadcast Channel," vol. 52, no. 9, pp. 3936–3964, Sept. 2006.

[154] D. P. Williamson and D. B. Shmoys, *The design of approximation algorithms.* Cambridge University Press 2011.

[155] Y. Wu, S. Shamai and S. Verdu, "Degrees of Freedom of the Interference Channel: a General Formula," in *Proc. IEEE Int. Symp. Inform. Theory,* Aug. 2011, pp. 1362–1366.

[156] L.-L. Xie and P. R. Kumar, "On the path-loss attenuation regime for positive cost and linear scaling of transport capacity in wireless networks," *IEEE Trans. Inform. Theory*, vol. 52, no. 6, pp. 2313–2328, Jun. 2006.

[157] F. Xue, L.-L. Xie, and P. R. Kumar, "The transport capacity of wireless networks over fading channels," *IEEE Trans. Inform. Theory*, vol. 51, no.3, pp. 834–847, Mar. 2005.

[158] A. C. Yao, "Some complexity questions related to distributive computing," in *Proc. of ACM STOC*, 1979, pp. 209–213.

[159] R. Yates and D. Tse, "Fading Broadcast Channels with State Information at the Receivers," Apr. 2009. [Online]. Available: `http://arxiv.org/abs/0904.3165`

[160] S. M. S. Yazdi and S. A. Savari, "A Max-Flow/Min-Cut Algorithm for Linear Deterministic Relay Networks," *IEEE Trans. Inform. Theory,* vol. 57, no. 5, pp. 3005–3015, 2011. Preliminary version in *Proc. of ACM-SIAM SODA*, 2010.

[161] L. Ying, R. Srikant, and G. E. Dullerud, "Distributed symmetric function computation in noisy wireless sensor networks," *IEEE Trans. Inform. Theory,* vol. 53, no. 12, pp. 4826–4833, Dec. 2007.